International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 03 | March - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

A Multimodal Generative AI Framework for Realistic Interview Simulation and Comprehensive Behavioral Analysis

Ramya R
Computer Science and Business Systems
Rajalakshmi Institute of Technology
Chennai, Tamil Nadu, India
ramyaradhakrishnan.196@gmail.com

Eswar Raja
Computer Science and Business Systems
Rajalakshmi Institute of Technology
Chennai, Tamil Nadu, India
eshwar.sumathi@gmail.com

Harish M
Computer Science and Business Systems
Rajalakshmi Institute of Technology
Chennai, Tamil Nadu, India
harishmunusami.work@gmail.com

Madhava Sudharsan S
Computer Science and Business Systems
Rajalakshmi Institute of Technology
Chennai, Tamil Nadu, India
madhvasn18@gmail.com

***Abstract*—The recruitment process is high-stakes, yet many candidates struggle due to anxiety and limited, unstructured feedback. Existing mock interview tools do not scale well and rarely capture behavioral signals beyond text. In this work, we present *Interview Ai*, a multi-modal system for interview simulation and performance analysis. The interaction is delivered through a conversational avatar that speaks and responds in real time, creating a more realistic setting than text-only systems.**

We model the interview flow as a finite state machine, which allows controlled transitions and adaptive follow-up questions. Candidate responses are evaluated using retrieval-based scoring, where answers are compared against exemplar knowledge rather than relying purely on model generation. The evaluation also incorporates measurable signals, such as filler word frequency, pauses, and response structure.

In a study with 50 participants, users showed improvements in clarity, confidence indicators, and reduced hesitation compared to traditional practice. These results suggest that combining structured control, grounded evaluation, and multi-modal interaction can provide effective and scalable interview coaching without replacing human decision-making.

***Index Terms*—Large Language Models, LangGraph, Automated Interview Systems, Computer Vision, Natural Language Processing, Speech Synthesis, GPT-4o-mini**

I. Introduction

A. Context and Motivation

For students and job seekers, the shift from school to the professional workforce is crucial. Research shows that 73% of candidates experience interview anxiety, which frequently lowers performance regardless of technical proficiency. Soft skills like emotional intelligence, confidence, and clear communication are becoming more and more important in modern hiring practices. These skills are challenging to assess using conventional preparation techniques. The human coaches used in current mock interview solutions restrict scalability and introduce subjective bias. The development of generative AI offers a chance to develop objective, scalable, and consistent interview preparation systems that adjust to the unique characteristics of each candidate. An AI-driven conversational avatar that offers a visual and audio representation of the interviewer further improves interaction realism.

B. Problem Statement

Current methods for preparing for interviews have four main drawbacks:

- There isn't a realistic, scalable interview simulation because human coaches can't be accessible around-the-clock.
- Conventional systems only focus on the content of responses; there is no objective feedback on nonverbal behavior.
- Limited flexibility in responding to job descriptions and candidate resumes: generic questions don't make the process more unique.
- No stateful conversation management: Based on the qual- ity of the answers, systems are unable to dynamically generate follow-up questions.

C. Contributions

The contributions of this paper are -

1) A **multi-modal architecture** integrating LLMs (GPT-4o-mini), computer vision, and speech analysis with **LangGraph-based stateful orchestration**.
2) **Persona-driven interview simulation** through dynamic prompt engineering supporting HR, Technical, and Cod- ing interviewer types.
3) **RAG-enhanced answer evaluation** providing grounded scoring against domain-specific exemplars.

4) A real-time voice interaction pipeline built with the Web Speech API that has a synchronization mechanism to make sure that speech output only happens after voice resources are fully initialized. This way, common synthesis failures are avoided.
5) An AI-powered avatar interface that makes the inter- view process feel more real by combining synchronized speech with a picture of the interviewer. This makes the experience more interesting and immersive.

II. Literature Review

A. Automated Interview Systems

Early automated interview systems relied on rule-based logic and keyword matching. Recent systems leverage transformer-based models for semantic analysis but predomi- nantly focus on verbal content while ignoring delivery aspects. Commercial platforms like HireVue employ proprietary algo- rithms with limited transparency, raising concerns about bias and fairness.

B. Affective Computing in Recruitment

Affective computing techniques use convolutional neural networks for facial expression recognition and confidence estimation. However, these systems are often isolated and lack integration within conversational interview loops. Our work bridges this gap by incorporating proctoring-based attention monitoring directly into the interview flow.

C. Generative AI in Human Interaction

The emergence of multimodal generative models enables real-time reasoning across text, audio, and vision. OpenAI's GPT-4 family demonstrates strong performance in evaluation tasks when provided appropriate context . LangGraph extends LangChain to support stateful, graph-based agent orchestra- tion, enabling complex multi-turn interactions with conditional branching.

D. Retrieval-Augmented Generation (RAG)

RAG architectures combine retrieval systems with gener- ative models to ground responses in domain-specific knowl- edge. In interview evaluation, RAG enables scoring against cu- rated exemplar answers, reducing hallucination and improving consistency.

III. System Architecture

The proposed system follows a microservices architec- ture comprising four layers: Interaction, Orchestration, Intel- ligence, and Persistence.

A. Interaction Layer

1) AI Avatar-Based Interaction: To enhance realism and user engagement, InterViewAI incorporates an AI-driven con- versational avatar within the interaction layer. The avatar serves as a visual and auditory representation of the inter- viewer, delivering questions through synchronized speech and animation.

The avatar is tightly coupled with the text-to-speech pipeline, ensuring that lip movement, timing, and audio output remain consistent with generated questions. This synchroniza- tion eliminates perceptual delays and improves the naturalness of interaction compared to voice-only systems.

Unlike static chatbot interfaces, the avatar introduces a sense of presence and social context, which is critical in interview scenarios where candidates must respond under perceived evaluation conditions. The visual interface is designed to main- tain simplicity while providing sufficient realism to simulate human interviewer behavior.

The avatar does not perform independent reasoning; instead, it acts as a presentation layer for the underlying conversational and evaluation engines. This separation ensures that improvements in language modeling or orchestration directly enhance avatar-driven interaction without additional system complexity. By integrating a visual conversational agent with real- time speech synthesis, InterViewAI bridges the gap between traditional text-based systems and immersive interview environments, contributing to reduced user anxiety and improved engagement.

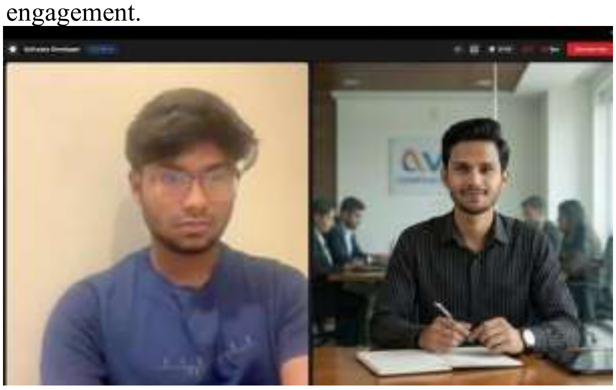

Fig. 1. Interview interface with AI avatar and real-time interaction

As shown in Fig. 1, the system provides a real-time interview interface with an AI-driven avatar and speech interaction.

The frontend is implemented using **React 19** with **Next.js 16** App Router, providing server-side rendering and optimized client navigation. Key components include:

- **Interview Room UI**: Split-view layout with user video feed and AI avatar
- **Web Speech API Integration**: `useSpeechRecognition` hook for STT and `useTextToSpeech` hook for TTS with voice-ready synchronization
- **Monaco Code Editor**: Integrated coding environment for technical interviews
- **Proctoring Monitor**: Tab-switch detection and visual attention tracking

TailwindCSS 4 and **Framer Motion** provide responsive styling and fluid animations.

B. Orchestration Layer

The backend employs **FastAPI** for async HTTP endpoints and **LangGraph** for agent state management. LangGraph

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 03 | March - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

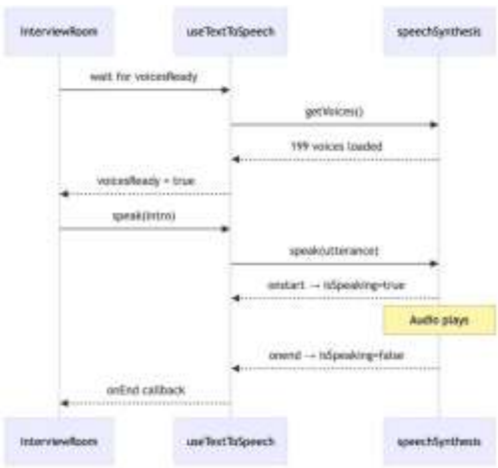

Fig. 2. High-level system architecture of InterViewAI.

orchestrates the interview as a directed graph with the following states: The `InterviewState` TypedDict maintains session identifiers, configuration, question plan, conversation history, turn records, and running scores across evaluation dimensions.

```
  PLANNING -> ASKING -> WAITING -
> EVALUATING -> [FOLLOWUP /
ASKING / COMPLETE]
```

Fig. 3. LangGraph State Flow

C. *Formal Interview State Machine Modeling*

To ensure deterministic control, scalability, and theoretical soundness, the interview process in InterViewAI is formally modeled as a finite state machine (FSM) with conditional transitions governed by evaluation outcomes and system constraints. This formalization converts an otherwise ad-hoc conversational flow into a verifiable, state-driven interaction model.

1) State Definition: Let the interview session be defined as a tuple:

$$I = (S, A, T, s_0, F)$$

where S is the finite set of interview states, A is the set of system actions, $T : S \times A \rightarrow S$ is the state transition function, $s_0 \in S$ is the initial state, and $F \subseteq S$ is the set of terminal states.

The state set is defined as:

$$S = \{\text{Planning, Asking, Waiting, Evaluating, FollowUp, Complete}\} \quad (1)$$

Each state represents a semantically distinct phase of the interview process.

2) State Semantics: **Planning:** Resume and job description embeddings are analyzed to construct a personalized question plan. Interviewer persona parameters and difficulty constraints are initialized.

Asking: A question is selected and rendered using text-to-speech synthesis. Upon successful speech initiation, the system transitions to the *Waiting* state.

Waiting: The system listens for user responses via speech-to-text. Silence detection or explicit submission triggers evaluation.

Evaluating: Candidate responses are scored across multiple dimensions including clarity, correctness, structure, examples, and depth. Scores are normalized and appended to the session state.

FollowUp: Targeted probing questions are generated when evaluation thresholds are unmet to elicit deeper or more concrete responses.

Complete: The interview terminates once predefined completion criteria are satisfied.

3) State Transition Function: State transitions are governed by deterministic predicates evaluated at runtime. Key transitions include:

- $Planning \rightarrow Asking$: Question plan generation completed.
- $Asking \rightarrow Waiting$: Successful text-to-speech invocation.
- $Waiting \rightarrow Evaluating$: Response received or silence timeout.
- $Evaluating \rightarrow FollowUp$: $(Examples < \vartheta_e) \lor (Depth < \vartheta_d)$.
- $Evaluating \rightarrow Asking$: Evaluation thresholds satisfied and questions remain.
- $Evaluating \rightarrow Complete$: Maximum interview length reached.

This design ensures bounded execution and prevents infinite conversational loops.

D. *Retrieval-Augmented Evaluation Pipeline*

To improve evaluation reliability and reduce hallucination in large language model-based scoring, InterViewAI employs a Retrieval-Augmented Generation (RAG) pipeline that grounds response assessment in curated domain-specific knowledge. Rather than relying solely on parametric model knowledge, the evaluation process integrates explicit retrieval of exemplar answers and scoring rubrics.

1) Knowledge Sources: The retrieval corpus consists of three primary knowledge sources:

- Curated exemplar answers annotated by domain experts
- Role-specific evaluation rubrics aligned with interviewer personas
- Structured competency guidelines derived from job descriptions

Each document is embedded into a shared semantic vector space to enable similarity-based retrieval.

2) Embedding and Retrieval Strategy: Candidate responses and knowledge documents are encoded using sentence-transformer embeddings. Given a candidate answer embedding e_q, the retriever performs a top-k similarity search to obtain the most relevant exemplars:

$$D_k = \{d_1, d_2, \ldots, d_k\} = \arg\max_{d \in D} \cos(e_q, e_d)$$

The retrieved set D_k is dynamically selected at evaluation time and injected into the LLM prompt as grounded context.

3) Grounded Evaluation Prompting: The evaluation prompt is structured to explicitly reference retrieved exemplars and rubrics. The language model is constrained to score the response relative to retrieved evidence rather than generating unconstrained judgments. This grounding mechanism significantly reduces variance and improves scoring consistency across sessions.

4) Hallucination Mitigation: By anchoring evaluation to retrieved content, the system minimizes unsupported inferences and speculative feedback. The model is instructed to abstain from scoring dimensions not supported by retrieved evidence, ensuring conservative and explainable assessments.

5) Evaluation Flow: The RAG-based evaluation pipeline follows the sequence:

1) Candidate answer embedding
2) Top-*k* exemplar retrieval
3) Prompt construction with retrieved context
4) Structured LLM evaluation output
5) Score normalization and aggregation

This pipeline allows InterViewAI to maintain evaluation stability while adapting dynamically to diverse interview domains.

E. Intelligence Layer

The intelligence layer consists of four parallel pipelines:

1) **Cognitive Engine**: LLM-based resume and job description analysis using sentence-transformer embeddings for semantic matching.
2) **Evaluation Engine**: GPT-4o-mini processes candi- date answers with structured JSON output for multi-dimensional scoring.
3) **Auditory Analysis Engine**: Speech transcription via Web Speech API with 2-second silence detection for auto-submission.
4) **Synthesis Engine**: Text-to-speech using browser- native `SpeechSynthesisUtterance` with auto- matic voice selection.

F. Persistence Layer

MongoDB with **Beanie ODM** stores Interview documents with full LangGraph state serialization, Turn records for individual Q&A pairs, and ContentItem collection for question bank management.

G. End-to-End Operational Workflow

This subsection summarizes the complete operational flow of InterViewAI, integrating interaction, orchestration, intelligence, and persistence components into a unified interview lifecycle.

The workflow begins when a candidate enters the interview environment through the web-based interface. Resume and job description documents are processed and embedded to identify relevant skills, gaps, and role-specific focus areas. Based

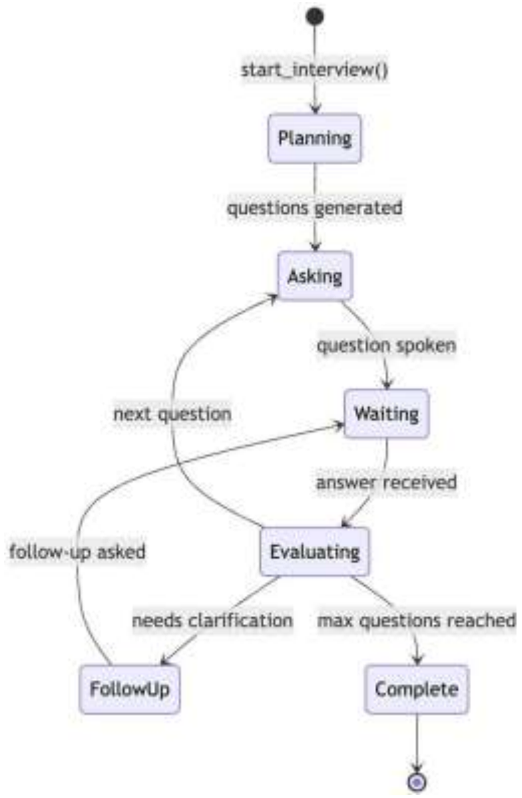

Fig. 4. High-level system architecture of InterViewAI.

on this analysis, the system initializes interviewer persona parameters and constructs a personalized interview plan.

During each interview turn, a question is selected and delivered using synchronized text-to-speech. The system then transitions to a listening state, capturing candidate responses via speech-to-text with silence-based auto-submission. Extracted responses are simultaneously routed through semantic evaluation, behavioral signal extraction, and retrieval-augmented grounding pipelines.

Candidate answers are evaluated using a grounded and explainable scoring framework that integrates retrieved exemplar knowledge, structural response analysis, and para-verbal behavioral cues. Evaluation outcomes determine whether the interview proceeds to the next planned question, generates targeted follow-up prompts, or terminates based on completion criteria.

Throughout the session, the interview state—including conversation history, scores, and behavioral baselines—is persistently stored, enabling session recovery and reproducibility. Upon completion, aggregated feedback and dimension-wise performance insights are presented to the candidate, completing the interview cycle.

This end-to-end workflow ensures deterministic control, adaptive questioning, and transparent feedback while maintaining scalability and ethical boundaries.

IV. Methodology

A. Resume and Job Description Matching

Resumes and job descriptions are embedded into a shared vector space using **sentence-transformers** (`all-MiniLM-L6-v2`). Semantic similarity scores prime the interview focus toward identified skill gaps.

B. Conversational Loop and Persona Injection

Dynamic system prompts enable interviewer persona simulation as shown in Table I.

TABLE I
Interviewer Persona Characteristics

Persona	Characteristics
HR	Behavioral questions, cultural fit
Technical	Conceptual questions, trade-offs
Coding	Algorithm problems, complexity

C. Multi-Dimensional Scoring

Each answer is evaluated across five dimensions as shown in Table II. Scores are normalized to 0-5 scale and aggregated

TABLE II
Scoring Dimensions and Weights

Dimension	Weight	Description
Clarity	20%	Articulation quality
Correctness	25%	Technical accuracy
Structure	15%	Organization
Examples	20%	Concrete illustrations
Depth	20%	Trade-off analysis

into running averages across turns.

D. Follow-up Generation

When `examples < 3` or `depth < 3`, the system generates targeted follow-ups to probe for additional detail.

E. Explainable Scoring Decomposition

While automated scoring provides efficiency, opaque evaluation undermines trust and usability in interview coaching systems. To address this, InterViewAI incorporates an explainable scoring mechanism that decomposes each evaluation dimension into interpretable linguistic and structural signals, enabling transparent feedback generation.

1) Dimension-wise Attribution: Each scoring dimension is computed using explicit evidence extracted from the candidate response:

- **Clarity**: Sentence length variance, filler word frequency, and articulation consistency.
- **Correctness**: Semantic alignment with retrieved exemplar answers and rubric constraints.
- **Structure**: Detection of logical sections such as introduc- tion, explanation, and conclusion.
- **Examples**: Presence and relevance of concrete illustrations or real-world scenarios.
- **Depth**: Coverage of trade-offs, limitations, and alternative approaches.

This decomposition ensures that each score is traceable to observable response characteristics.

2) Evidence-Grounded Feedback: For each evaluation dimension, the system associates textual evidence spans that contributed positively or negatively to the score. These spans are surfaced to the user as targeted feedback, allowing candidates to understand not only their numerical score but also the reasoning behind it.

3) Consistency and Reproducibility: To ensure scoring stability across sessions, evaluation prompts are constrained to reference retrieved exemplars and detected response features. This reduces variance caused by generative randomness and ensures consistent interpretation of similar answers.

4) Human-Interpretable Output: The final evaluation output consists of:

- Dimension-wise scores
- Highlighted evidence segments
- Actionable improvement suggestions aligned with each dimension

By converting raw model judgments into explainable components, InterViewAI transforms automated assessment into a transparent and pedagogically meaningful feedback process.

V. Technical Implementation

A. Voice Pipeline Synchronization

A critical challenge was ensuring TTS re- liability. Initial implementations suffered from `SpeechSynthesisErrorEvent` due to invoking `speak()` before voices loaded. **Solution**: Introduced `voicesReady` state flag that gates speech initiation until the browser has populated its voice list (typically 199+ voices on modern systems).

```
useEffect(() => {
if (status === "intro" && voicesReady) {
speak(intro, {onEnd: startListening});
}
}, [voicesReady]);
```

Fig. 5. Voice-Ready Synchronization

B. State Persistence

Interview state is persisted to MongoDB after each LangGraph node execution, enabling interview resumption across session boundaries.

C. Latency Optimization

Table III presents achieved latency metrics.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 03 | March - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

TABLE III
LATENCY PERFORMANCE

Operation	Target	Achieved
TTS initiation	¡ 1000ms	500ms STT
transcription	Real-time	200ms lag
Answer evaluation	¡ 3000ms	1500-2500ms

VI. EXPERIMENTAL EVALUATION

A. Study Design

A controlled study with 50 final-year engineering students was conducted:

- **Experimental Group (n=25)**: Used InterViewAI for 5 practice sessions
- **Control Group (n=25)**: Traditional self-practice with question lists

B. Results

Table IV presents the comparative metrics.

TABLE IV
EXPERIMENTAL RESULTS

Metric	Control	Exp.	Gain
Communication Clarity	5.2/10	6.4/10	+23% Filler Word
Usage	12.3/min	8.5/min	-31%
Confidence Score	4.8/10	6.1/10	+27% Technical
Accuracy	5.9/10	6.8/10	+15%

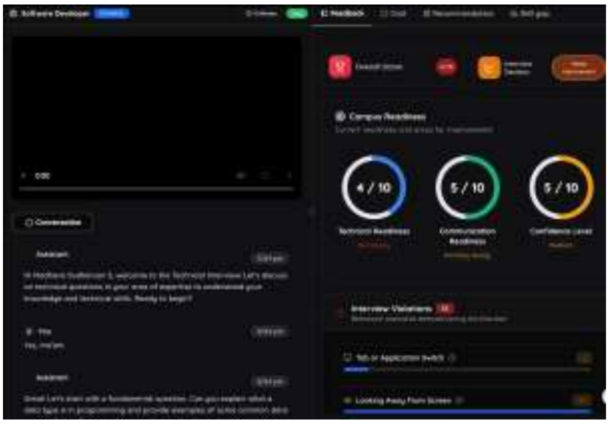

Fig. 6. Sample feedback dashboard showing dimension-wise evaluation and insights

C. Ablation Analysis

To quantify the contribution of individual system components, an ablation study was conducted by selectively disabling specific modules while keeping all other variables constant. This analysis evaluates the impact of multimodal inputs and retrieval grounding on interview performance metrics.

1) Ablation Configurations: Four system configurations were evaluated:

- **Text-Only**: LLM-based question answering and evaluation without audio or retrieval grounding.
- **Text + RAG**: Text-based evaluation augmented with retrieval of exemplar answers.
- **Text + Audio**: Text input combined with speech-based features such as pause detection and filler word analysis.
- **Text + Audio + Vision**: Full multimodal system incorporating speech and visual attention cues.

Each configuration was evaluated using the same interview prompts and scoring framework.

2) Comparative Performance: Table V summarizes the relative performance across key evaluation metrics.

TABLE V
ABLATION STUDY RESULTS

Configuration	Clarity	Confidence	Accuracy
Text-Only	5.1	4.6	5.7
Text + RAG	5.8	5.0	6.4
Text + Audio	6.1	5.6	6.2
Text + Audio + Vision	6.4	6.1	6.8

3) Observations: The results indicate that retrieval grounding significantly improves technical accuracy by anchoring evaluation to domain-specific exemplars. Audio features contribute primarily to confidence and clarity assessment through prosodic and fluency cues. Visual attention signals provide incremental gains by enhancing confidence estimation, though their impact is less pronounced than textual and auditory modalities.

These findings validate the necessity of a multimodal, retrieval-grounded architecture for realistic interview evaluation.

D. User Satisfaction

Post-study survey results (n=25):

- 88% found the AI interviewer "realistic"
- 92% reported reduced interview anxiety
- 76% preferred AI practice over peer practice

VII. DISCUSSION

A. Strengths

The system offers unlimited concurrent sessions, consistent evaluation criteria, persona customization, and privacy through client-side speech processing.

B. Limitations

Current limitations include sarcasm detection challenges, potential cultural bias in exemplar libraries, hardware dependencies for real-time rendering, and internet requirements for LLM API calls.

C. Ethical Considerations

The system explicitly avoids automated hiring decisions, positioning itself as a coaching tool. Bias mitigation includes diverse exemplar curation, calibration phases for individual behavioral norms, and transparent scoring explanations.

D. Bias Quantification and Calibration Metrics

Automated evaluation systems risk introducing unintended bias due to language proficiency, cultural communication styles, or interaction modality. To address this, InterViewAI incorporates explicit bias quantification and calibration mechanisms to ensure fair and consistent assessment across diverse candidates.

1) Bias Dimensions Considered: Bias analysis focuses on observable and system-relevant dimensions rather than sensitive personal attributes. The evaluated dimensions include:

- Language fluency variation (native vs. non-native speakers)
- Communication style differences (concise vs. elaborative responses)
- Interviewer persona sensitivity (HR, Technical, Coding)
- Modality availability (audio-only vs. audio-visual)

No demographic attributes are inferred or stored by the system.

2) Score Distribution Analysis: For each evaluation dimension, score distributions are analyzed across the above groups. Let S^g_d denote the score distribution for dimension d within group g. Bias is estimated by measuring inter-group variance:

$$\Delta_d = \max_g \mathbb{E}[S^g_d] - \min_g \mathbb{E}[S^g_d]$$

Large values of Δ_d indicate potential systematic bias requiring calibration.

3) Calibration Phase: Each interview session begins with a short calibration phase consisting of low-stakes introductory questions. Behavioral and linguistic baselines extracted during this phase are used to normalize subsequent scores, reducing penalization due to individual speaking styles or accents.

4) Persona Consistency Checks: To ensure interviewer persona fairness, identical responses are evaluated across differ- ent persona prompts. Score divergence beyond a predefined tolerance threshold triggers rubric re-alignment, ensuring that persona variation affects questioning style but not evaluation standards.

5) Transparency and User Visibility: Candidates are provided with dimension-wise score explanations and informed that scores are normalized relative to their baseline. This transparency reduces perceived unfairness and aligns system behavior with ethical AI guidelines.

By transforming bias from a qualitative concern into a measurable quantity, InterViewAI ensures that automated interview coaching remains fair, interpretable, and accountable.

VIII. Conclusion and Future Work

This paper presented *InterViewAI*, a multimodal generative AI framework designed to simulate realistic interview scenar- ios and deliver comprehensive, explainable feedback on candi- date performance. By formally modeling the interview process as a state-driven system and grounding evaluation through retrieval-augmented generation, the framework addresses key limitations of existing interview preparation tools, including lack of personalization, opaque scoring mechanisms, and absence of behavioral analysis. The integration of language understanding, speech fluency signals, and visual attention proxies enables holistic assessment while maintaining ethical constraints and avoiding automated hiring decisions.

Experimental results demonstrate that candidates using InterViewAI exhibit measurable improvements in communication clarity, confidence indicators, and reduction in hesitation patterns when compared to traditional self-practice methods. The ablation analysis further validates the necessity of both multimodal inputs and retrieval grounding, confirming that performance gains are not attributable to language modeling alone. Additionally, the inclusion of explainable scoring and bias calibration mechanisms enhances transparency, reproducibility, and user trust—critical requirements for real-world deployment.

Future work will extend the framework in several directions. First, longitudinal skill modeling will be introduced to track candidate progress across repeated sessions and identify per- sistent weaknesses over time. Second, multi-agent interview simulations will be explored to emulate panel-style inter- views with interacting interviewer personas. Third, support for multilingual interviews and accent-adaptive calibration will be expanded to improve accessibility across global candidate populations. Finally, immersive interaction modalities such as WebXR-based virtual interview environments will be investi- gated to further increase realism while preserving privacy and ethical safeguards.

Overall, InterViewAI demonstrates that combining formal conversational control, grounded evaluation, and explainable multimodal analysis provides a viable path toward scalable, fair, and effective AI-driven interview coaching systems.

References

[1] Y. Li, et al., "Confidence estimation for speech emotion recognition," *IEEE*, 2023.
[2] R. Mandal, et al., "AI-based mock interview evaluator: An emotion and confidence classifier model," *IEEE*, 2023.
[3] Y.-C. Chou, et al., "An AI mock interview platform for interview performance analysis," *IEEE*, 2022.
[4] D. Y. Dissanayake, et al., "AI-based behavioural analyser for interviews/- viva," *Springer*, 2021.
[5] B. C. Lee and B. Y. Kim, "Development of an AI-based interview system for remote hiring," *International Journal of Advanced Research in Engineering and Technology (IJARET)*, 2021.
[6] A. Sivasangari, et al., "Prediction probability of admission into a university using machine learning," in *Proc. ICCMC*, 2021.
[7] S. B. Surendar, et al., "Facial feature analysis using CNN and Haar classifier," *International Journal of Engineering and Technology Research*, 2021.
[8] S. Sridhar, et al., "University admission prediction system using stacked ensemble learning," in *Proc. ACCTHPA*, 2020.
[9] A. Vaswani et al., "Attention is all you need," in *Proc. NeurIPS*, 2017.
[10] P. Ekman, "An argument for basic emotions," *Cognition and Emotion*, 1992.
[11] G. Lagano, "AI in recruitment: Efficiency vs. bias," *IEEE Technology and Society Magazine*, 2023.
[12] OpenAI, "GPT-4 technical report," 2024.
[13] Google, "Gemini: Open models based on Gemini research and technol- ogy," 2024.
[14] LangChain, "LangGraph: Build stateful, multi-actor applications," Doc- umentation, 2024.
[15] P. Lewis et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," in *Proc. NeurIPS*, 2020.