Vol. 1, Issue 1 | Jan 2026 | Open Access — Peer Reviewed Research Paper

AI-Based Interview Preparation Platform

Vedant N. Shinde

Department of Computer Science and Engineering

Agasti Institute of Management Computer Application and Research (AIMCAR), Akole

Abstract—Effective interview preparation requires repeated, realistic practice and specific feedback on both verbal content and delivery, resources that most candidates cannot access on demand due to the cost and scheduling difficulty of arranging mock interviews with a qualified human reviewer. This paper presents the design, implementation, and evaluation of an AI-Based Interview Preparation Platform that simulates a role-specific interview, evaluates spoken responses across content and delivery dimensions, and generates targeted feedback to guide subsequent practice. The system integrates four components — a question generation module that selects and adapts questions from a role-tagged question bank using retrieval-augmented prompting, a speech-to-text and prosody analysis pipeline that transcribes responses and extracts delivery metrics such as pace, filler-word frequency, and pause distribution, a response evaluation module that scores transcript content for relevance, structure, and completeness using a fine-tuned sentence-embedding similarity model, and a feedback generation module that converts the combined scores into specific, actionable written feedback. A corpus of 3,400 role-specific interview questions across six job families and 1,150 candidate response recordings with expert-assigned reference scores was used to train and validate the evaluation components. The content scoring module achieved a Pearson correlation of 0.86 with expert human ratings, and the delivery metrics extraction pipeline achieved a word error rate of 6.4% on held-out audio. A four-week pilot with 35 final-year engineering students showed a statistically significant improvement in self-reported interview confidence and a measurable reduction in filler-word frequency across successive practice sessions. The paper documents the system architecture, module design, evaluation methodology, observed limitations relating to accent and domain coverage, and the future scope of the work.

Index Terms—Interview Preparation, Natural Language Processing, Speech Recognition, Automated Feedback, Sentence Embeddings, Prosody Analysis, Mock Interview Simulation, Career Readiness, Educational Technology, Response Evaluation, Machine Learning, Decision Support System.

I. INTRODUCTION

Interview performance depends on two largely independent skills: the content of a response, which reflects domain knowledge and structured reasoning, and the delivery of that response, which reflects pace, clarity, and confidence. Improving either skill requires repeated practice under realistic conditions followed by specific corrective feedback, a combination that is difficult for most candidates to access outside a small circle of mentors, career counsellors, or paid coaching services [1], [2].

Placement cycles in higher education compress this preparation need into a short window during which large numbers of students require practice simultaneously, far exceeding the capacity of available human interviewers. Self-directed alternatives such as reading question banks or rehearsing alone in front of a mirror provide practice without feedback, leaving candidates unable to identify which specific habits are undermining their performance [3].

Recent advances in speech recognition and natural language understanding make it feasible to automate a meaningful portion of this feedback loop. A system can transcribe a spoken response, measure delivery characteristics

B. Question Bank and Generation Module

A curated bank of role-tagged questions spans technical, behavioural, and situational categories across six job families. The generation module selects an initial question sequence based on the candidate's stated role and experience level, and adapts follow-up questions using retrieval over semantically related items when a response signals a particular area of strength or weakness.

C. Response Capture Module

Candidate responses are recorded through the browser microphone and streamed to a speech-to-text service. The resulting transcript is time-aligned with the audio to support downstream extraction of pace and pause metrics at the word level.

D. Delivery Analysis Pipeline

The delivery pipeline computes speaking rate in words per minute, filler-word frequency per hundred words, and the distribution of pause durations between phrases. These metrics are normalised against role-appropriate reference ranges derived from the expert-rated response corpus to produce a delivery sub-score.

directly from the audio signal, and compare the transcript's content against a reference answer structure using sentence-embedding similarity, producing feedback that is specific rather than generic [4].

This paper proposes an AI-Based Interview Preparation Platform that simulates a role-specific interview session, evaluates each response across both content and delivery dimensions, and returns feedback that a candidate can act on before the next practice attempt.

The principal contributions of this work are: (1) a retrieval-augmented question generation module that adapts a role-tagged question bank to the candidate's stated job family and experience level; (2) a delivery analysis pipeline that extracts pace, filler-word frequency, and pause metrics from spoken responses; (3) a content evaluation module that scores transcript relevance, structure, and completeness against reference answers; (4) a feedback generation module that converts these scores into specific written guidance; and (5) an empirical evaluation covering scoring accuracy and a four-week candidate pilot.

II. RELATED WORK

A. Automated Interview Assessment

Early automated interview systems relied on keyword matching against a fixed answer key, which penalised candidates who expressed correct content in unanticipated phrasing. Studies of such systems report low correlation with expert ratings, particularly for open-ended behavioural questions [5], [6].

B. Speech and Prosody Analysis

Automatic speech recognition combined with prosodic feature extraction has been used to quantify delivery characteristics such as speaking rate, pitch variation, and disfluency frequency. Research in this area consistently finds that filler-word frequency and pause patterns correlate with perceived candidate confidence, motivating their inclusion as explicit scored metrics [7], [8].

C. Sentence Embedding for Content Evaluation

Transformer-based sentence embeddings allow semantic similarity to be measured between a candidate transcript and a reference answer without requiring exact lexical overlap, substantially improving correlation with human graders relative to keyword-based scoring on short-answer evaluation tasks [9], [10].

D. Feedback Generation for Skill Training

Research on automated feedback in skill-training contexts, including public speaking and language learning, finds that specific, dimension-linked feedback produces greater improvement across repeated practice sessions than an

E. Content Evaluation Module

Each transcript is encoded using a fine-tuned sentence-embedding model and compared against a reference answer structure for the corresponding question, scoring relevance, structural organisation, and completeness. Responses falling below a calibrated confidence threshold are flagged for optional review by a human mentor where one is available.

F. Feedback Generation Module

The feedback module combines the content and delivery sub-scores with a template-guided natural language generation step that produces specific observations, such as identifying an under-developed section of a response or a recurring filler-word pattern, together with a suggested focus for the next practice attempt.

G. Progress Dashboard

A dashboard tracks each candidate's content and delivery sub-scores across successive sessions, visualising trend lines per question category so that improvement or persistent weakness becomes visible over time rather than only within a single session.

IV. METHODOLOGY

A. Dataset Preparation

A corpus of 3,400 role-specific interview questions was assembled across six job families, and 1,150 candidate response recordings were collected with expert-assigned reference scores for content relevance, structure, completeness, and delivery quality, each on a five-point scale.

B. Experimental Protocol

The labelled response corpus was partitioned into training, validation, and test subsets in a 70:15:15 ratio. The content scoring model's similarity threshold and the delivery pipeline's normalisation parameters were tuned on the validation subset, and all reported metrics were computed on the held-out test subset.

C. Baseline Models

Three content scoring approaches were compared: keyword overlap scoring representing conventional systems, TF-IDF cosine similarity, and the fine-tuned sentence-embedding model used in the final system. All approaches were evaluated against the same expert reference scores.

D. Evaluation Metrics

Content scoring quality was measured using Pearson correlation with expert human ratings. Delivery pipeline quality was measured using word error rate on held-out audio and mean absolute error between extracted and manually annotated pace and pause metrics.

aggregate score alone, a finding that directly informs the feedback design adopted in this work [11], [12].

III. SYSTEM ARCHITECTURE

A. Architectural Overview

The proposed system follows a four-layer architecture consisting of a question layer, a capture layer, an evaluation layer, and a feedback layer, as summarised in Fig. 1. Layer separation allows the speech pipeline and the content scoring model to be updated independently of the question bank and the user interface.

E. Implementation Environment

The prototype was implemented in Python 3.11 using a cloud speech-to-text service for transcription, sentence-transformers for content embedding, and Flask for the application layer. The interface was built with HTML5, CSS3, and Chart.js for the progress dashboard, and all experiments were executed on a machine with an Intel Core i5 processor and 8 GB of memory.

F. Pilot Study Design

A four-week pilot was conducted with 35 final-year engineering students preparing for campus placements. Participants completed one practice session per week, with self-reported interview confidence and filler-word frequency tracked across sessions to assess change over the pilot period.

Fig. 1 — Layered System Architecture of the Proposed AI-Based Interview Preparation Platform

Layer	Principal Components and Responsibilities
Feedback Layer	Feedback generation, progress dashboard, trend visualisation, mentor escalation
Evaluation Layer	Content scoring model, delivery analysis pipeline, threshold calibration
Capture Layer	Browser microphone capture, speech-to-text service, transcript time-alignment
Question Layer	Role-tagged question bank, retrieval-based follow-up question generation

TABLE I — Comparative Analysis: Self-Practice vs. Static Question Bank App vs. Proposed System

Feature	Self-Practice Alone	Static Question Bank App	Proposed AI Platform	Benefit to Candidate
Question Selection	Manual browsing	Fixed list by category	Role-adaptive, follow-up aware	Relevant, targeted practice
Content Feedback	None	None	Embedding-based relevance score	Specific content gaps identified
Delivery Feedback	Self-assessment only	Not available	Pace, filler-word, pause analysis	Objective delivery metrics
Feedback Detail	Not available	Generic tips	Dimension-linked written feedback	Actionable next steps
Progress Tracking	None	None	Session-over-session dashboard	Visible improvement trend
Availability	Depends on partner	Always available	Always available	Practice on demand
Scalability	Not scalable	Scalable, low value	Scalable, high value	Suitable for large cohorts

TABLE II — Model Performance, Pipeline Accuracy, and Pilot Study Results

#	Model / Module	Metric	Result	Observation
1	Keyword overlap scoring	Pearson r	0.52	Penalises valid rephrasing
2	TF-IDF cosine similarity	Pearson r	0.68	Improves on keyword matching
3	Sentence-embedding model (selected)	Pearson r	0.86	Closest to expert agreement
4	Speech-to-text pipeline	Word error rate	6.4%	Held-out audio test set

#	Model / Module	Metric	Result	Observation
5	Pace / pause extraction	Mean absolute error	4.1%	Against manual annotation
6	Pilot — filler-word reduction	Session 1 vs. 4	31.2%	Averaged across 35 participants

V. RESULTS AND DISCUSSION

A. Content Scoring Accuracy

As reported in Table II, the fine-tuned sentence-embedding model achieved a Pearson correlation of 0.86 with expert human ratings, substantially exceeding keyword overlap scoring at 0.52 and TF-IDF cosine similarity at 0.68. The improvement was most pronounced for behavioural questions, where candidates frequently express equivalent content using varied phrasing that lexical methods cannot capture.

Residual disagreement between the model and expert raters was concentrated in responses that were topically relevant but poorly structured, a dimension the current similarity-based approach captures only indirectly through reference-answer structure rather than through an explicit discourse-level signal.

B. Delivery Pipeline Accuracy

The speech-to-text pipeline achieved a word error rate of 6.4% on held-out audio, and extracted pace and pause metrics recorded a mean absolute error of 4.1% against manual annotation. Accuracy was highest for clear, moderate-pace speech and lowest for responses with heavy accent variation or background noise, both common in uncontrolled recording environments.

C. Pilot Study Outcomes

Across the four-week pilot, average filler-word frequency fell by 31.2% between the first and fourth practice sessions, and self-reported interview confidence, measured on a five-point scale, rose from an average of 2.8 to 3.9. Participants identified the dimension-linked written feedback and the session-over-session trend view as the two most useful features.

D. Discussion

The results support the central premise of this work: separating content and delivery evaluation, and returning specific rather than generic feedback on each, allows candidates to identify and correct particular habits across successive practice attempts. The magnitude of the pilot improvements should be interpreted cautiously, since the cohort was drawn from a single institution and self-reported confidence is inherently subjective.

VI. LIMITATIONS

Several limitations qualify these findings. First, the speech-to-text pipeline's accuracy degrades for candidates with strong regional accents or non-native pronunciation patterns

Adding video-based non-verbal analysis, including eye contact and posture estimation, would extend feedback beyond the audio-and-text dimensions currently covered. Expanding question bank coverage to additional job families and industry domains would improve relevance for specialised candidates. Finally, a multi-institution pilot would strengthen the evidence base for the platform's effect on interview readiness.

VIII. CONCLUSION

This paper presented an AI-Based Interview Preparation Platform that integrates role-adaptive question generation, delivery analysis, content evaluation, and dimension-linked feedback generation within a single practice pipeline. The content scoring module achieved a Pearson correlation of 0.86 with expert human ratings, and the delivery pipeline achieved a word error rate of 6.4% on held-out audio.

A four-week pilot with 35 candidates demonstrated a 31.2% reduction in filler-word frequency and a measurable increase in self-reported interview confidence, indicating that specific, dimension-linked feedback can drive behavioural improvement across successive practice sessions. The limitations relating to accent coverage, factual verification, and pilot scale define a clear agenda for the extensions outlined above. The work shows that combining automated speech analysis with embedding-based content evaluation can convert isolated self-practice into a structured, feedback-driven preparation process.

IX. REFERENCES

[1] All India Council for Technical Education, Employability Skills Assessment Report, AICTE, New Delhi, 2024.

[2] Sharma, N. and Iyer, V., Barriers to Effective Interview Preparation Among Engineering Graduates, Journal of Career Development Studies, vol. 11, pp. 66–81, 2023.

[3] Deshpande, M., Self-Directed Practice and Feedback Deficiency in Placement Preparation, Indian Journal of Educational Technology, vol. 7, pp. 29–44, 2022.

[4] Vaswani, A. et al., Attention Is All You Need, Advances in Neural Information Processing Systems, vol. 30, pp. 5998–6008, 2017.

[5] Hoque, M. et al., MACH: My Automated Conversation Coach, Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 697–706, 2013.

underrepresented in the training corpus, which can distort delivery metrics for affected users. Second, the question bank currently covers six job families, so candidates preparing for niche or highly specialised roles receive less targeted question coverage.

Third, the content scoring model evaluates relevance and structure against a reference answer but does not verify factual correctness, so a confidently stated but technically incorrect response can still receive a favourable content score. Fourth, the four-week pilot involved 35 participants from a single institution and cannot support claims about generalisation across diverse educational backgrounds or industries. Finally, the platform does not currently model non-verbal cues such as eye contact or posture, which are known to influence interview outcomes but require video-based analysis beyond the present audio-and-text pipeline.

VII. FUTURE SCOPE

Future development will proceed along five directions. Expanding the training corpus with accented and non-native speech samples would improve delivery pipeline robustness across a wider candidate population. Incorporating a lightweight fact-verification step against domain knowledge bases would allow the content scorer to flag technically incorrect responses rather than scoring them on relevance alone.

[6] Naim, I. et al., Automated Analysis and Prediction of Job Interview Performance, IEEE Transactions on Affective Computing, vol. 9, pp. 191–204, 2018.

[7] Nguyen, L. et al., Hire Me: Computational Inference of Hirability in Employment Interviews, IEEE Transactions on Multimedia, vol. 16, pp. 1018–1031, 2014.

[8] Pentland, A., Honest Signals: How They Shape Our World, MIT Press, Cambridge, 2008.

[9] Reimers, N. and Gurevych, I., Sentence-BERT: Sentence Embeddings Using Siamese BERT-Networks, Proceedings of EMNLP-IJCNLP, pp. 3982–3992, 2019.

[10] Devlin, J., Chang, M., Lee, K. and Toutanova, K., BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, Proceedings of NAACL-HLT, pp. 4171–4186, 2019.

[11] Zhao, R. et al., Automated Public Speaking Feedback with Virtual Reality, Proceedings of the ACM Conference on Computer-Supported Cooperative Work, pp. 1825–1837, 2017.

[12] Chen, L. and Yoon, S., Application of Multimodal Feedback in Language Skill Training, Computer Assisted Language Learning, vol. 32, pp. 604–626, 2019.

[13] National Association of Software and Service Companies, Campus Readiness and Employability Trends Report, NASSCOM, 2024.

[14] Ministry of Education, Higher Education and Employability Survey 2024, Government of India, New Delhi, 2024.

[15] Shinde, V. and Gaikwad, P., Pilot Evaluation of an AI-Assisted Mock Interview Platform, National Conference on Emerging Computing Technologies, 2026 (in press).

© 2026 IJVAST — International Journal of Vast Advancements in Science and Technology.
ISSN: Applied For (2026) | DOI: 10.XXXXX/IJVAST.2026.XXXXX | Open Access — CC BY 4.0 License