Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

Smart AI Interviewer and Resume Analyzer

Pooja Vachkal
Dept. of Computer Engg.
JSCOE, Hadapsar Pune, India

Vishal Chole
Dept. of Computer Engg.
JSCOE, Hadapsar Pune, India

Gaurav Padol
Dept. of Computer Engg.
JSCOE, Hadapsar Pune, India

Samarth Kawane Dept. of Computer Engg.
JSCOE, Hadapsar Pune, India

Omkar Kasar
Dept. of Computer Engg.
JSCOE, Hadapsar Pune, India

Abstract—With job markets growing increasingly competitive, candidates require structured and objective tools to evaluate their interview readiness. This paper presents IndusAI, a web-based intelligent coaching system that automatically assesses interview performance through multimodal analysis of speech, language, and non-verbal behavior. The platform employs Whisper-based transcription, transformer NLP models, and prosody extraction to generate a quantified confidence score alongside personalized, actionable feedback. An integrated resume evaluation engine examines document structure and keyword alignment to estimate Applicant Tracking System (ATS) screening likelihood. Upon completing analysis, the system compiles a downloadable PDF report that consolidates all findings and targeted improvement recommendations, offering candidates a data-driven pathway to close the gap between academic preparation and industry hiring standards.

Keywords—AI Interview Coaching, Resume ATS Evaluation, Automatic Speech Recognition, Whisper ASR, BERT, Sentence-BERT, Confidence Scoring, NLP, Prosody Analysis, MediaPipe, Career Readiness, Job Placement.

I. INTRODUCTI

Securing employment in a technically demanding field requires candidates to demonstrate strong verbal communication, domain knowledge, and professional presentation skills. Despite this, a large proportion of job seekers exit interviews without any structured insight into how they performed, making targeted self-improvement difficult.

An equally critical yet frequently overlooked challenge is resume screening. Modern recruitment pipelines rely on Applicant Tracking Systems (ATS) to shortlist candidates automatically based on keyword relevance and formatting compliance. Resumes that do not satisfy these criteria are discarded before a human hiring manager ever reviews them, placing technically qualified candidates at an undeserved disadvantage.

IndusAI addresses both challenges within a single, accessible web platform. The system conducts AI-driven mock interview sessions, evaluates candidate responses across verbal, acoustic, and semantic dimensions, and simultaneously performs ATS-oriented resume assessment. All analytical outputs are consolidated into a structured PDF coaching report that candidates can download and act upon immediately.

The proposed solution integrates Whisper for high-accuracy speech transcription, BERT and Sentence-BERT for semantic content evaluation, openSMILE for voice prosody feature extraction, and MediaPipe for optional non-verbal posture and gaze analysis. This multimodal architecture enables a holistic, quantified assessment of candidate job-readiness that no single-modality tool can replicate.

Because IndusAI operates entirely through a standard web browser with no specialized hardware requirements, it is equally suitable for individual students, university placement cells, and corporate pre-screening workflows across diverse industries and geographies.

II. LITERATURE REV

Scholarly interest in machine-assisted interview evaluation and automated resume analysis has expanded considerably in recent years, fueled by rapid progress in large-scale pre-trained language models, multimodal perception frameworks, and speech recognition technology. The following works form the primary intellectual foundation of IndusAI.

Nagasawa et al. (2024) [IEEE Trans. Affective Computing] investigated the design of interview robots capable of adapting their questioning strategy in real time according to a candidate's detected speaking willingness. Their findings underscore the value of continuous affective monitoring during interviews, a principle that shapes IndusAI's dynamic confidence scoring mechanism.

Artiran et al. (2022) [IEEE Trans. Neural Syst. Rehabil. Eng.] examined gaze behavior differences in individuals with Autism Spectrum Condition during immersive VR interview scenarios. The gaze and head-orientation metrics introduced in this work provided a methodological basis for IndusAI's non-verbal analysis module, implemented via MediaPipe landmark tracking.

Ashrafi et al. (2023) [IEEE Access] developed a resume-driven re-education framework for career transitions in fast-evolving labor markets. Their keyword relevance scoring and job fit matching strategies served as an architectural reference for the ATS compliance and resume relevance components within IndusAI.

Stoev et al. (2025) [IEEE Access] evaluated BERT-derived embeddings for the automated classification of adult attachment interview transcripts in German, achieving classification accuracy comparable to expert human evaluators. This outcome validated the adoption of transformer-based representations for IndusAI's semantic

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

content scoring pipeline.

Radford et al. (2023) [Proc. ICML] released Whisper, a sequence-to-sequence transformer pre-trained on approximately 680,000 hours of weakly supervised multilingual speech data. Its word-level timestamp capability enables IndusAI to precisely locate and quantify filler-word occurrences and compute per-segment speech rate without manual annotation.

Reimers & Gurevych (2019) [Proc. EMNLP] introduced Sentence-BERT, adapting the BERT architecture with Siamese and triplet network training objectives to produce semantically meaningful fixed-length sentence vectors. IndusAI uses cosine similarity between these embeddings to measure how closely a candidate's answer aligns with expert reference responses.

III. PROPOSED METHODOLOGY

IndusAI is architected as a six-stage processing pipeline. An interview recording together with an optional resume document serves as the system input; the pipeline terminates with a unified Confidence Score and an auto-compiled PDF coaching report delivered directly to the candidate.

A. Input Module

The ingestion stage accepts audio-visual recordings in MP4, WAV, and MP3 formats alongside resume files in PDF or DOCX. Embedded audio tracks are extracted from video containers and passed to the acoustic pipeline, while document files are routed to the resume assessment engine independently.

B. Speech and Acoustic Analysis

Whisper ASR converts the interview audio into a full transcript annotated with word-level timestamps. openSMILE subsequently extracts a 384-dimensional prosodic and spectral feature vector capturing pitch variance, energy contour, and speaking tempo. Filler-word density and syllable-per-second rate are derived from the aligned transcript and contribute to the voice sub-score.

C. Semantic and Linguistic Analysis

A fine-tuned BERT classifier assigns sentiment polarity and grammatical quality ratings to each candidate turn. Sentence-BERT then encodes both the candidate response and a domain-specific reference answer into dense vectors; cosine distance between these vectors quantifies semantic relevance and forms the content sub-score.

D. Non-Verbal Behavioral Analysis (Optional)

When a video stream is available, MediaPipe's face mesh and pose estimation models extract per-frame gaze vectors and upper-body keypoints. Temporal aggregation yields eye-contact ratio and postural stability indices, which jointly constitute the non-verbal sub-score.

E. Resume ATS Assessment

The resume parser extracts section structure, heading labels, and body tokens. A keyword matching engine compares extracted tokens against a configurable target job-description vocabulary, reporting a keyword coverage ratio and formatting compliance score that predict ATS screening outcome.

F. Score Aggregation and Report Generation

All feature vectors are min-max normalized to [0, 1]. The composite score is computed as: $C = 100 \times (w1*Sv + w2*Sf + w3*Sc + w4*Snv - P)$, where $w1+w2+w3+w4 = 1$ and P denotes disfluency and compliance penalties. Thresholds classify output as **Highly Competent** ($C \geq 75$), **Satisfactory** ($50 \leq C < 75$), or **Requires Development** ($C < 50$). A structured PDF report presenting scores, trend charts, and improvement guidance is rendered on demand.

IV. SYSTEM ARCHITECTURE & FEATURES

The IndusAI architecture is organized into six loosely coupled functional layers. Each layer handles a distinct responsibility, allowing independent testing, easy component upgrades, and horizontal scaling without affecting other parts of the system.

A. User Interface Layer

A server-rendered web application exposes upload forms for interview media and resume documents, a session dashboard for tracking historical assessments, and a report viewer with inline score breakdowns. The layout is fully responsive and conforms to standard web accessibility guidelines.

B. Media Ingestion Layer

An asynchronous file handling service validates MIME types, extracts raw audio from video containers using FFmpeg, and routes each data stream to the appropriate processing worker. Supported input types include MP4, WebM, WAV, MP3, PDF, and DOCX.

C. Multimodal Analytics Layer

Four concurrent workers handle speech transcription (Whisper), prosody extraction (openSMILE), semantic scoring (BERT/Sentence-BERT), and non-verbal analysis (MediaPipe). Worker outputs are normalized and collected by a score aggregation service before proceeding to the next layer.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

D. Confidence Scoring Lay

A configurable weight matrix combines the four sub-scores and applies rule-based deduction penalties for excessive disfluency, ATS non-compliance, and sustained gaze avoidance. The final scalar score and categorical label are persisted to the session store.

E. Report Synthesis Lay

A template engine populates a structured PDF layout with score summaries, per-section radar charts, highlighted resume gaps, and NLP-generated improvement recommendations tailored to each candidate's weak points.

F. Extensibility and Sca

The stateless worker architecture supports horizontal scaling via container orchestration. Planned extensions include multilingual resume parsing, integration with live job-board APIs for real-time keyword benchmarking, and an institutional analytics dashboard for cohort-level readiness monitoring.

V. GAP ANALYSIS

Table I contrasts IndusAI against conventional interview preparation approaches. Human-led mock sessions, while valuable, are constrained by cost, scheduling availability, and evaluator subjectivity. Existing mobile feedback applications address at most one or two assessment dimensions. IndusAI uniquely unifies speech, language, affect, non-verbal, and resume evaluation within a single automated pipeline accessible without institutional affiliation.

TABLE I
COMPARATIVE ASSESSMENT: CONVENTIONAL METHODS VS. INDUSAI

Criterion	Conventional Approach	IndusAI
Feedback Source	Human evaluator	Automated AI pipeline
Speech Scoring	Absent / informal	Whisper ASR + openSMILE
Resume Review	Manual recruiter check	ATS keyword simulation
Body Language	Subjective observation	MediaPipe gaze & posture
Output Report	Verbal / handwritten	Structured PDF download
Accessibility	Appointment-based only	24/7 on-demand web app
Scalability	Bounded by staff capacity	Unlimited via cloud deploy

VI. RESU

The deployed IndusAI web application was exercised across a diverse set of interview recordings and resume documents to verify end-to-end pipeline correctness and interface usability. Figures 1 and 2 capture key screens of the operational system. Fig. 1 depicts the platform landing page, which presents the core value proposition alongside entry points for voice evaluation, real-time expression tracking, AI-driven question generation, and session smart reports. Fig. 2 illustrates the video upload workflow, guiding users through file selection and confirming the four automated processing stages: audio extraction, speech-to-text conversion, NLP keyword identification, and question set generation.

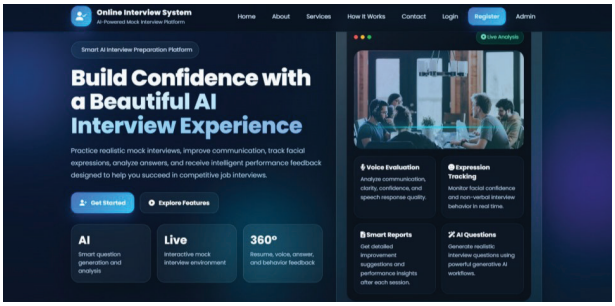

Fig. 1. Platform landing page illustrating the AI-powered interview preparation interface.

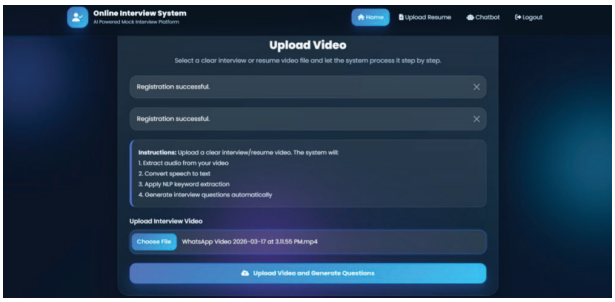

Fig. 2. Interview video upload screen demonstrating the automated processing workflow.

VII. CONCLU

This paper has introduced IndusAI, a multimodal AI platform that reframes interview preparation as a data-driven, self-directed activity. By fusing large-scale speech recognition, transformer-based language understanding, acoustic prosody modeling, optional non-verbal tracking, and ATS-aware resume assessment, the system generates a comprehensive, objective snapshot of candidate job-readiness in a single session.

Unlike conventional coaching approaches that depend on human availability and carry inherent evaluator bias, IndusAI delivers consistent, reproducible feedback at any time and at minimal cost per session. The platform is therefore well positioned to democratize high-quality interview coaching for students and professionals who lack access to expensive career services, regardless of their geographic location or institutional affiliation.

Future enhancements include adaptive question generation personalized to candidate performance trajectories, multilingual resume parsing for non-English

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

job markets, live job-portal API integration for dynamic keyword benchmarking, and a cohort analytics module enabling institutional placement coordinators to monitor collective readiness trends across graduating batches.

ACKNOWLEDGMENT

The authors express sincere appreciation to Pooja Vachkal for sustained mentorship and technical guidance throughout this project. Gratitude is also extended to the faculty and staff of the Department of Computer Engineering, Jayawantrao Sawant College of Engineering, Hadapsar, Pune (Savitribai Phule Pune University, Academic Year 2025–26), whose institutional support was instrumental in bringing IndusAI to fruition.

REFERENCES

[1] F. Nagasawa, S. Okada, T. Ishihara, and K. Nitta, "Adaptive Interv Strategy Based on Interviewees' Speaking Willingness Recognition for Interview Robots," *IEEE Trans. Affective Comput.*, vol. 15, no. 2, pp. 230–242, Feb. 2024.

[2] S. Artiran, R. Ravisankar, S. Luo, L. Chukoskie, and P. Cosm "Measuring Social Modulation of Gaze in Autism Spectrum Condition With Virtual Reality Interviews," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 2373–2385, Sept. 2022.

[3] S. Artiran, P. S. Bedmutha, and P. Cosman, "Analysis of Gaze, H Orientation, and Joint Attention in Autism With Triadic VR Interviews," *Frontiers Virtual Reality*, vol. 5, pp. 1–13, Mar. 2023.

[4] S. Ashrafi, B. Majidi, E. Akhtarkavan, and S. H. R. Hajiag "Efficient Resume-Based Re-Education for Career Recommendation in Rapidly Evolving Job Markets," *IEEE Access*, vol. 11, pp. 124350–124367, Nov. 2023.

[5] T. Stoev, E. Flemming, B. Strauss, K. Petrowski, C. Spitzer, and Yordanova, "Towards Automated Classification of Adult Attachment Interviews Using BERT," *IEEE Access*, vol. 13, pp. 155305–155320, Sept. 2025.

[6] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, and Sutskever, "Robust Speech Recognition via Large-Scale Weak Supervision," in *Proc. ICML*, 2023, pp. 28492–28518.

[7] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BE Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proc. NAACL*, 2019, pp. 4171–4186.

[8] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence Embeddi Using Siamese BERT-Networks," in *Proc. EMNLP*, 2019, pp. 3982–3992.

[9] C. Lugaresi et al., "MediaPipe: A Framework for Building Percept Pipelines," *arXiv*, 2019, arXiv:1906.08172.

[10] F. Eyben, M. Wöllmer, and B. Schuller, "openSMILE – The Mu Versatile and Fast Open-Source Audio Feature Extractor," in *Proc. ACM MM*, 2010, pp. 1459–1462.