

# AI-Driven Interview Simulation with Voice, Vision Analysis and Resume Evaluation

Rupali Anantrao Zamare<sup>1</sup>, Shubham Dhakare<sup>2</sup>, Saurabh Dhurve<sup>3</sup>, Veneth Bansod<sup>4</sup>, Raj Dudhe<sup>5</sup>, Shekh Shoyeb<sup>6</sup>

<sup>1</sup>Assistant Professor, Student, Department of Computer Science and Engineering, Shri Sant Gajanan Maharaj College of Engineering, Shegaon, Maharashtra, India

<sup>2,3,4,5,6</sup>Student, Department of Computer Science and Engineering, Shri Sant Gajanan Maharaj College of Engineering, Shegaon, Maharashtra, India

DOI: 10.5281/zenodo.23436381

## ABSTRACT

*Interview preparation has been shifting away from static question banks toward AI systems that can actually reason about the candidate, the role they're applying for, and how the interview itself unfolds. This paper describes an AI-driven interview simulation framework that brings together resume evaluation, adaptive question generation, automatic speech recognition, voice-feature analysis, computer-vision analysis, and LLM-based answer assessment. The system starts by pulling structured competencies out of a resume and comparing them to a job description using semantic representations. From there, it generates role-specific technical and behavioral questions and runs a multi-turn spoken interview. Throughout the session, speech is transcribed and scored for relevance, correctness, structure, fluency, speaking rate, pauses, and other delivery traits that can actually be observed. Video is analyzed for nonverbal cues — face visibility, gaze direction, head movement, and rough posture stability. A multimodal fusion layer then produces a scorecard that's tied back to specific evidence from what the candidate said. Recent work has shown that integrated virtual interviewers, multimodal candidate assessment, and LLM-based resume matching are all technically feasible, but it has also made clear that automated scores shouldn't be treated as a stand-in for human judgment [1]–[5]. With that in mind, the system described here leans toward coaching candidates rather than automating hiring — it prioritizes transparent evidence, honest uncertainty, privacy, fairness, and keeping a human in the loop.*

**Keywords** — AI interview, mock interview, resume matching, speech analysis, computer vision, multimodal learning, NLP, large language models, employability assessment, explainable AI.

## 1. INTRODUCTION

Recruitment interviews ask candidates to demonstrate technical knowledge, problem-solving, communication, and behavioral competence, often all in the same 30-minute conversation. Traditional mock interviews are useful practice, but they're hard to scale, and they tend to keep resume review, question generation, speech coaching, and behavioral feedback in separate silos. Recent AI systems suggest these pieces don't have to stay separate. Bershika and Nancy, for instance, built a web-based virtual interviewer that combines resume parsing, semantic matching, adaptive question generation, speech analysis, and facial-expression analysis into one pipeline [1]. PolyInterview takes a similar approach, pairing a target job description and CV with multi-turn spoken interviewing and multimodal assessment [2].

This work is aimed squarely at interview preparation, not at automating recruitment decisions. Its main contribution is a unified pipeline where resume evidence actually shapes which questions get asked, where answers are analyzed across multiple modalities, and where the resulting feedback is anchored to concrete evidence rather than an opaque score. That creates a closed learning loop: spot a competency gap, ask a question that targets it, measure the response, give feedback, and suggest another round of practice.

Three problems needed solving to get there. First, resume evaluation has to go beyond keyword overlap toward genuine semantic alignment between candidate and job. Second, interview assessment has to combine content and delivery signals without treating noisy behavioral cues as if they were hard proof of competence. Third, the whole multimodal pipeline needs to stay interpretable and privacy-aware. These concerns aren't abstract — an observational study of 736 resumes found only a weak correlation between GPT-4 ratings and human ratings, which is a fairly direct warning that LLM-based scores can't simply substitute for human assessment [3].

## 2. RELATED WORK

### 2.1. AI Interviewers and Adaptive Questioning.

LLMs have made it practical to generate context-aware questions and follow-ups on the fly. Several recent simulated-interview platforms draw on resumes, job descriptions, retrieval mechanisms, and LLM agents to build interview conversations that feel personalized [2], [6]. The advantage over a fixed question bank is obvious once you've used one: the system can push back when an answer is thin, or dig into a specific project mentioned on the resume instead of asking something generic.

### 2.2. Multimodal Interview Assessment.

Agrawal et al. proposed multimodal behavioral analytics that combine facial expressions, speech, prosody, and transcripts to give feedback on engagement, speaking rate, and eye contact [7]. Earlier still, Asada et al. looked at facial expression and verbal response during interview answers, using video, eye- and mouth-area features, response latency, and fundamental voice frequency [8]. More recently, researchers have started folding computer vision and LLM-based verbal analysis into assessment constructs that are actually meant to be interpretable [4].

### 2.3. Resume Matching and NLP.

Resume-job matching has moved on from simple keyword methods toward contextual embeddings and LLMs. Vaishampayan et al. compared GPT-4 ratings against human ratings across 736 resumes and found that prompt engineering could improve the quality of LLM ratings — but they were careful to note that LLM and human ratings still aren't interchangeable [3]. Modern semantic matching can represent skills, experience, education, and projects at the sentence or section level, which produces a much more meaningful match than simply checking whether the same words appear in both documents.

### 2.4. Speech and Vision Models.

Speech analysis typically runs automatic speech recognition first, then layers linguistic and acoustic analysis on top. Whisper showed strong multilingual, robust ASR performance and is a solid choice as the recognition component [9]. Self-supervised representations like wav2vec 2.0 give you rich acoustic features for whatever downstream task you're building [10]. On the vision side, OpenFace handles facial behavior analysis — landmarks, head pose, gaze, facial action units — reasonably well [11]. These tools are good at surfacing observable communication signals, but they shouldn't be pushed further than that; they're not equipped to infer someone's hidden mental state, and treating them that way would be a mistake.

### 2.5. Ethics and Applicant Reactions.

AI-supported recruitment shapes how candidates experience fairness, transparency, and basic interpersonal treatment during the process. Research on applicant reactions has found that people respond negatively when they perceive the AI as opaque or beyond their control [12]. There's also a growing body of critical scholarship questioning whether facial-analysis systems can validly infer internal traits from appearance or micro-expressions at all [13]. Given that, the system described here deliberately limits its vision output to observable communication indicators and frames the whole tool as a coach, not a judge.

## 3. PROPOSED SYSTEM ARCHITECTURE

The system is organized into six layers: (1) candidate/job ingestion, (2) resume evaluation, (3) adaptive interview planning, (4) real-time voice and vision capture, (5) modality-specific assessment, and (6) explainable report generation. The candidate uploads a resume and picks a target job. That job description gets converted into a set of competency requirements, while the resume parser pulls out education, experience, projects, certifications, skills, and tools. Semantic matching then links the evidence in the resume to what the role actually requires.

### 3.1. Resume Evaluation.

Let  $S$  be semantic similarity between resume sections and job requirements,  $K$  be weighted skill overlap,  $E$  be experience alignment, and  $P$  be project relevance. Resume-role compatibility is scored as  $R = w_1S + w_2K + w_3E + w_4P$ , with configurable weights that sum to one. Every score comes with a piece of supporting evidence attached — something like "Python skill supported by Project X" or "cloud experience not found in resume." That evidence trail matters; a single unexplained number isn't very useful to a candidate trying to improve.

### 3.2. Adaptive Interview Engine.

The engine builds an interview plan out of technical, project, behavioral, and situational questions, each one tagged with a competency, a difficulty level, and the kind of evidence it's meant to draw out. After every answer, the coverage vector gets updated and a follow-up is chosen accordingly — a shallow project explanation might trigger a debugging question, while a strong answer can bump the candidate into a tougher

design question. PolyInterview has already shown that this kind of answer-aware follow-up questioning, grounded in the job and CV, is workable in practice [2].

### 3.3. Voice Analysis.

The microphone stream gets converted to text via ASR, and the transcript is scored for relevance, technical coverage, coherence, and completeness. On the acoustic side, the system tracks things like words-per-minute, pause duration, speech-to-silence ratio, pitch variation, and energy. These are meant purely as communication-coaching signals, not as some proxy for personality. Noise level, microphone quality, accent, and ASR confidence are all logged as uncertainty factors rather than folded silently into the score.

### 3.4. Vision Analysis.

Selected video frames get processed for face presence, rough gaze direction, head orientation, facial movement, and posture stability. What the system reports are observable events — frequent gaze deviation, or the face disappearing from frame for an extended stretch. It stops there deliberately: it never claims that a facial expression proves honesty, confidence, intelligence, or general suitability for a job. That caution is backed up by recent multimodal-assessment research, which supports interpretable constructs while also flagging real risks in nonverbal inference [4], [13].

### 3.5. Multimodal Fusion.

For a given response, a content score  $C$ , voice score  $V$ , vision score  $G$ , and resume-role evidence score  $R$  are combined as  $F = \alpha C + \beta V + \gamma G + \delta R$ . Content carries the most weight in technical interviews, and vision carries the least — it can be dropped entirely when video quality isn't good enough to trust. The final report always shows the score alongside confidence and quality indicators, rather than presenting a bare number.

## 4. IMPLEMENTATION METHODOLOGY

A reasonable prototype would pair a React (or similar) browser frontend with a Python/Node.js backend. The browser side handles consent, camera and microphone access, the interview display, and candidate controls. The backend does the heavier lifting: resume extraction, embedding generation, question planning, ASR, audio-feature extraction, vision processing, LLM evaluation, and report generation. A vector database can hold job and competency embeddings, while a relational database stores the structured interview metadata.

Resume processing starts with text extraction and normalization, followed by section segmentation to identify education, work experience, projects, and skills. A controlled skill vocabulary handles synonyms and abbreviations so "ML" and "machine learning" don't get treated as unrelated. Sentence-level embeddings support section-to-requirement similarity, and the system keeps track of the exact evidence spans behind every important match, so nothing important is a black box.

During the interview itself, audio is transcribed close to real time. The LLM gets the current question, the transcript, the target competency, and the relevant resume evidence, and returns rubric scores for relevance, correctness, completeness, reasoning, and structure. The audio module works independently to produce delivery indicators, and the vision module processes sampled frames on its own track. The fusion layer then pulls all of this together and writes a short, evidence-linked explanation.

Any real evaluation protocol should track both system-level and user-level measures. ASR performance is naturally measured with word error rate. Resume matching can be checked against precision, recall, ranking correlation, and expert agreement. Answer scoring should be validated against expert rubrics using correlation, mean absolute error, and inter-rater reliability. On the user side, a study should look at perceived usefulness, feedback clarity, confidence before and after practice, and whether performance actually improves across repeated sessions.

To keep model drift and hallucination in check, generated questions should stay constrained to the selected job and the resume evidence at hand, prompts should be retrieval-augmented, the rubric should stay explicit, and unsupported claims should be rejected outright. Reports should always separate what was actually detected from what the model is interpreting.

## 5. EXPECTED RESULTS AND DISCUSSION

The hope is that an integrated system like this delivers more actionable feedback than a text-only mock interview ever could. A candidate should walk away with both content feedback — missing technical concepts, say — and delivery feedback, like too many pauses or an answer that never really found its structure. That combination isn't just wishful thinking; recent studies back up its feasibility. PolyInterview reported strong alignment between its generated question plans and the matched job descriptions, along with expert-rated feedback that people actually found actionable [2]. A 2026 multimodal employability framework goes further still, folding textual resumes, video interviews, and audio responses into a single interpretable model [5].

One useful experimental setup would compare three configurations: Model A (resume plus text answers), Model B (resume plus text and audio), and Model C (resume plus text, audio, and vision). The study should try to

determine whether the extra modalities genuinely improve the usefulness of the feedback, or whether they mostly just introduce new reliability and fairness headaches. A separate experiment could compare keyword matching, sentence-embedding matching, and LLM-assisted matching against one another.

The metric that actually matters here is improvement in interview readiness — not the ability to predict hiring outcomes. Candidates should be able to revisit the same competency after getting feedback and show a measurable improvement the second time around. As an example, an answer that initially scores low for structure could be reattempted using the STAR framework, and the system can then check whether situation, task, action, and result all come through more clearly.

Automated scoring needs to stay advisory, not authoritative. Vaishampayan et al.'s resume study is a good reminder of why: GPT-4 and human ratings simply weren't interchangeable [3]. Similarly, research on speech-emotion systems notes that automated emotion recognition often needs human intervention, since the models frequently miss context that a person would catch [14]. That's why the scorecard here is built to show confidence, evidence, and known limitations alongside the score — not to present the AI's judgment as objective fact.

## **6. PRIVACY, FAIRNESS AND SECURITY**

An interview system like this handles genuinely sensitive personal data, so privacy has to be designed in from the start: explicit consent, minimal data collection, encryption in transit and at rest, restricted access, defined retention periods, and real deletion controls. Where possible, raw video and audio should be processed transiently and thrown away right after feature extraction — there's no good reason to keep the raw recording around longer than that.

Fairness has to be checked across accents, languages, genders, disabilities, lighting conditions, camera quality, and different communication styles. Candidates shouldn't be penalized for an accent or for disability-related speech patterns, full stop. Vision features should never stand alone as evidence of employability. There should also be a "video unavailable" path so candidates can still complete a content-focused practice session without any visual analysis at all.

On the security side, the system needs authentication, authorization, secure API endpoints, prompt-injection defenses, input sanitization, encrypted storage, and audit logs. Resumes and job descriptions are untrusted input and should never be allowed to override system instructions, no matter how the content is phrased. Voice recordings deserve extra protection too, since speech alone can reveal sensitive characteristics about the speaker [14].

Finally, transparency with the candidate isn't optional. Before the interview starts, the interface should spell out which modalities are being used, what's actually being measured, how the scores get calculated, and whether any data is being stored. That's not just good practice — it lines up with evidence that opaque AI recruitment tends to leave a bad taste with applicants [12].

## **7. CONCLUSION AND FUTURE WORK**

This paper laid out an AI-driven interview simulation framework that unifies resume evaluation, adaptive questioning, speech analysis, computer vision, and LLM-based answer assessment into one coaching platform. It uses the candidate's resume and target job to shape relevant questions, analyzes spoken responses, picks out observable communication patterns, and turns all of that into evidence-linked recommendations rather than a black-box grade.

The real research challenge going forward is building multimodal assessment that's actually reliable without overstating what AI can infer from human behavior — that line is easy to cross and hard to walk back. Future work should validate the framework against real interview datasets, compare different LLMs and ASR models against each other, add support for multilingual interviews, calibrate scores against expert rubrics, and study fairness across accents and demographic groups in more depth. And for anything resembling a high-stakes recruitment decision, human review needs to stay mandatory, not optional.

## **8. DETAILED ALGORITHMIC WORKFLOW**

The full workflow can be described as a sequence of evidence-preserving transformations. Stage 1 converts the resume and job description into structured text and competency representations. Stage 2 has the question planner pick a competency and pull the resume evidence that supports it. Stage 3 captures and transcribes the spoken answer. Stage 4 scores language, audio, and visual features independently of one another. Stage 5 has the fusion engine combine only the quality-controlled features and produce an evidence-linked explanation. Keeping these stages separate matters — a failure in one modality shouldn't be allowed to quietly invalidate the whole interview.

Three lightweight algorithms carry the AI modules: semantic embeddings, LLM-based question generation, and quality-aware multimodal fusion.

#### **ALGORITHM 1: RESUME-JOB SEMANTIC MATCHING**

Input: Resume R, Job Description J

1. Extract resume sections SR and job requirements RJ.
2. Generate embeddings ER and EJ for comparable sections.
3. Compute cosine similarities  $S_{ij}$  between candidate and job evidence.
4. Compute weighted skill, experience, and project scores K, E, P.
5. Return  $R_{score} = wS + xK + yE + zP$  with evidence spans.

#### **ALGORITHM 2: ADAPTIVE INTERVIEW QUESTION GENERATION**

Input: Competency state C, Resume R, History H, Job J

1. Select the highest-priority uncovered competency c.
2. Retrieve resume evidence supporting c.
3. Generate a question q constrained by J, R, and interview level.
4. After answer a, update coverage(c) and difficulty using rubric score.
5. Generate a follow-up only when evidence is incomplete or ambiguous.

#### **ALGORITHM 3: QUALITY-AWARE MULTIMODAL SCORING**

Input: Text T, Audio A, Vision V, Resume score R

1. Compute content score C from relevance, correctness, and completeness.
2. Compute voice score  $V_c$  only when audio/ASR quality exceeds threshold.
3. Compute vision score G only when frame quality and face visibility are valid.
4. Fuse  $F = \alpha C + \beta V_c + \gamma G + \delta R$  using interview-specific weights.
5. Report F with modality confidence and evidence-linked feedback.
1. Extract sections  $SR = \{\text{skills, education, experience, projects}\}$ .

A key design principle throughout is graceful degradation. If the camera isn't available, the system just continues with resume, text, and audio evidence. If the audio signal is poor, the transcript can still be kept while the acoustic scores get suppressed. If ASR confidence drops too low, the response gets flagged for manual correction rather than scored aggressively as if nothing were wrong. This kind of modality-aware behavior keeps the system reliable and stops technical glitches from turning into unfair penalties for the candidate.

### **9. SCORING RUBRIC AND EXPLAINABILITY**

The answer rubric covers five dimensions: relevance, technical correctness, completeness, reasoning quality, and communication structure. Each is scored on a five-point scale with clear descriptors attached — a 5 in technical correctness means the answer had accurate concepts and an appropriate explanation, while a 2 signals partial knowledge or a real gap. The rubric itself is shown to the candidate, either before or after practice, so the scoring never feels arbitrary.

Explainability works on two levels. Local explanations cover why one particular response got the score it did; global explanations summarize how the candidate did across the whole interview. A local report might note that an answer correctly explained model training but skipped validation entirely. A global report might say machine-learning fundamentals are solid but system-design reasoning still needs work. Wherever possible, evidence is linked directly to transcript spans or resume sections rather than left as an abstract claim.

Uncertainty is reported separately from performance, and that separation matters. An answer with genuinely strong content but low ASR confidence shouldn't get marked down just because the transcription itself was shaky. In the same way, a vision score always comes with image-quality and face-detection reliability attached. This follows a broader principle worth holding onto in any reliable AI system: how confident you are in a measurement and what that measurement is actually telling you are two different things, and they shouldn't get collapsed into one number.

### **10. EXPERIMENTAL DESIGN**

A controlled evaluation could run three groups of candidates through different feedback configurations. Group A gets text-only questions with content-only scoring. Group B gets text and audio feedback. Group C gets the full multimodal system. Each participant would complete an initial interview, a feedback session, and then a second interview a week later, with improvement measured at the competency level rather than as one overall score.

The main quantitative metrics would be expert-score correlation, mean absolute error, ranking consistency, ASR word error rate, and resume-to-job matching agreement. Secondary metrics could include response latency, question repetition rate, feedback acceptance, and system response time. On the user-experience side, Likert-

scale items can capture usefulness, clarity, trust, fairness, and willingness to try the simulation again, while qualitative interviews can surface cases where the automated feedback felt off or beside the point.

A useful ablation study would remove one modality at a time — comparing the full model against text-only, text-plus-audio, and text-plus-vision versions shows whether each modality is actually pulling its weight. A second ablation could compare keyword matching, embedding similarity, and LLM-assisted matching directly. Statistical tests should report effect sizes and confidence intervals rather than leaning entirely on p-values, and inter-rater agreement among human evaluators gives a practical upper bound to measure the automated scoring against.

The dataset itself should span a range of roles — software development, data science, cybersecurity, testing, general engineering — with each sample including the job description, an anonymized resume, the interview question, the candidate's response, and expert rubric annotations. Consent and de-identification aren't optional when real recordings are involved. Synthetic data is fine for early development, but the final evaluation really needs realistic accents, speaking rates, lighting conditions, and resume layouts to mean anything.

## 11. USE CASES AND DEPLOYMENT MODEL

The most obvious use case is individual interview prep: a student picks a target role, uploads a resume, and works through a 15–30 minute simulation, coming away with a report that flags role-specific gaps and suggests what to practice next. A second use case is placement-cell coaching, where instructors set up role profiles and assign mock interviews across a cohort, with the platform surfacing anonymized class-level trends without exposing anyone's individual recordings.

A third use case is professional upskilling — someone changing domains can use the system to spot missing terminology, weak project explanations, and communication gaps they didn't know they had. Organizations could also use it for internal interview training, as long as it's never deployed as an unreviewed hiring gate. In every case, the candidate stays the primary user and should have direct access to the evidence behind whatever feedback they're given.

For deployment, a hybrid edge-cloud design makes the most sense. Sensitive audio/video preprocessing can happen on the client or a controlled local service, while the less sensitive embeddings and rubric calculations run on a secure server. A configurable retention policy can wipe raw media immediately after feature extraction and keep only aggregate scores and whatever notes the candidate has approved. That setup cuts down on both privacy exposure and storage costs.

## 12. LIMITATIONS

A few limitations are worth naming plainly. Semantic matching can be thrown off by resumes that are incomplete or, on the other end, exaggerated. LLM-based answer scoring can be sensitive to prompt wording and to underlying model changes over time. ASR errors tend to hit accents, background noise, and code-switching harder than other speech patterns. Visual indicators are highly sensitive to camera position, lighting, and individual communication style — two equally strong candidates can look very different on camera for reasons that have nothing to do with their answers. And a clean numerical score can create a false sense of precision if the underlying rubric or evidence quality isn't actually solid.

These limitations point toward a fairly conservative deployment policy: prioritize content and job-relevant evidence, flag uncertainty rather than hide it, let candidates correct their own transcripts, and keep a human-review option available. Vision and emotion-related features should stay optional in coaching contexts, and they should never be used on their own to make a call about someone's employability.

## 13. REFERENCES

- [1] J. M. Bershika and G. Nancy, "AI-Based Virtual Interviewer System Using NLP and Emotion Detection," in *Sustainable Computing and Intelligent Systems (SCIS 2025)*, Lecture Notes in Networks and Systems, vol. 1929, pp. 69–85, Springer, 2026, doi: 10.1007/978-3-032-22911-3\_6.
- [2] Z. Wen, J. Cao, Z. Wang, C. Chen, X. Liu, J. Yin, and Z. Li, "PolyInterview: An LLM-based Platform for Immersive Mock Interview Practice with Comprehensive Multimodal Assessment," *arXiv preprint arXiv:2607.10310*, 2026.
- [3] S. Vaishampayan, H. Leary, Y. B. Alebachew, L. Hickman, B. Stevenor, W. Beck, and C. Brown, "Human and LLM-Based Resume Matching: An Observational Study," *Findings of NAACL 2025*, pp. 4823–4838, 2025, doi: 10.18653/v1/2025.findings-naacl.270.
- [4] K. Kassab, A. Kashevnik, and I. Shoshina, "Multi-Modal Method for Candidate Interview Assessment Based on Computer Vision and Large Language Models," *Big Data and Cognitive Computing*, vol. 10, no. 4, art. 106, 2026, doi: 10.3390/bdcc10040106.
- [5] F. Zhou, "Employability Assessment Using Multimodal Deep Learning Framework," *Discover Artificial Intelligence*, vol. 6, art. 348, 2026, doi: 10.1007/s44163-026-01041-5.

- [6] T. T. H. Nguyen, T. D. Q. Nguyen, H. L. Cao, T. C. T. Tran, T. C. M. Truong, and H. Cao, "SimInterview: Transforming Business Education through Large Language Model-Based Simulated Multilingual Interview Training System," arXiv preprint arXiv:2508.11873, 2025.
- [7] A. Agrawal, R. A. George, S. S. Ravi, S. Kamath, and A. Kumar, "Leveraging Multimodal Behavioral Analytics for Automated Job Interview Performance Assessment and Feedback," in Proc. Challenge-HML, pp. 46–54, 2020.
- [8] T. Asada, Y. Yoshitomi, and M. Tabuse, "A System for Analyzing Facial Expression and Verbal Response of a Person While Answering Interview Questions on Video," Journal of Robotics, Networking and Artificial Life, vol. 4, pp. 114–117, 2017, doi: 10.2991/jrnl.2017.4.2.2.
- [9] A. Radford et al., "Robust Speech Recognition via Large-Scale Weak Supervision," in Proc. ICML, PMLR, vol. 202, pp. 28492–28518, 2023.
- [10] A. Baevski, Y. Zhou, A. Mohamed, and M. Auli, "wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations," in Advances in Neural Information Processing Systems, vol. 33, 2020.
- [11] T. Baltrusaitis, P. Robinson, and L.-P. Morency, "OpenFace: An Open Source Facial Behavior Analysis Toolkit," in Proc. IEEE Winter Conf. Applications of Computer Vision (WACV), pp. 1–10, 2016, doi: 10.1109/WACV.2016.7477553.
- [12] T. Langer, J. König, and C. Papathanasiou, "Can I show my skills? Affective responses to artificial intelligence in the recruitment process," Review of Managerial Science, 2022, doi: 10.1007/s11846-021-00514-4.
- [13] C. "Facial analysis AI as social pseudotechnology," European Journal for Philosophy of Science, 2026, doi: 10.1007/s13194-026-00732-1.
- [14] "Speech emotion recognition systems and their security aspects," Artificial Intelligence Review, 2024, doi: 10.1007/s10462-024-10760-z.
- [15] N. Srinadh Reddy and C. S. Mrudhula, "An AI-Driven Multimodal Framework for Interview Readiness Assessment and Skill Enhancement," International Journal for Research in Applied Science & Engineering Technology, 2026, doi: 10.22214/ijraset.2026.79237.