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

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

AI-Based Hybrid Interview Readiness Assessment System Using NLP and Behavioral Feature Analysis

Amrutha S K
Dept. Artificial Intelligence and Data Science
BGS College of Engineering and Technology Be ngaluru, India

Priyanka M
Dept. Artificial Intelligence and Data Science
BGS College of Engineering and Technology Be ngaluru, India

Nisarga B
Dept. Artificial Intelligence and Data Science
BGS College of Engineering and Technology Bengaluru, India

Nithyashree T R
Dept. Artificial Intelligence and Data Science
BGS College of Engineering and Technology Bengaluru, India

Dr. Chandini A G
Dept. Artificial Intelligence and Data Science
BGS College of Engineering and Technology Bengaluru, India

Abstract - Remote technical assessment and automated candidate evaluation have emerged as critical components of contemporary recruitment and educational skill measurement. Conventional proctoring systems encounter significant operational challenges including elevated processing latency arising from server-side video transmission overhead, inflexible keyword-dependent resume screening mechanisms, and binary evaluation labels that discourage candidate motivation during practice. This paper presents Pranishree, a comprehensive client-side edge-computed multimodal assessment and proctoring platform that executes entirely within the browser environment without server-side video processing. The platform incorporates real-time acoustic speech processing through the Web Speech API, a multi-model edge computer vision engine comprising TensorFlow.js COCO-SSD, MediaPipe 3D FaceMesh, and Chrome Native FaceDetector, an 8,000-question repository spanning 100 specialized technical roles with uniform Fisher-Yates O(n) shuffling, and a domain-adaptive conversational AI resume coaching module. A synchronous warning management architecture using warningsRef.current enforces automatic session termination upon accumulation of four or more malpractice events including mobile phone detection, multi-person presence, candidate absence, and tab-switching. Experimental evaluation across 150 simulated interview sessions demonstrates high detection accuracy with F1 scores of 98.0% for multi-person detection, 98.8% for candidate absence, and 96.0% for mobile phone hardware identification, alongside sub-second processing latency not exceeding 350ms. Candidates completing three or more practice sessions recorded an average Interview Readiness Score improvement of 18% across consecutive attempts, demonstrating the platform's effectiveness in maintaining assessment security while sustaining candidate motivation.

I. INTRODUCTION

A. Background and Domain Context

The widespread transition toward remote engineering recruitment and automated educational assessment has generated substantial demand for scalable, objective, and secure evaluation platforms. Conventional manual technical interviews are resource-intensive, constrained by scheduling limitations, and prone to evaluator subjectivity. Automated multiple-choice assessments, while scalable, are insufficient for measuring verbal articulation, architectural reasoning, or problem-solving methodology—competencies that directly relate to on-the-job performance.

B. Operational Challenges

Three primary challenges characterize existing remote evaluation platforms. First, high latency and bandwidth consumption: many video proctoring systems transmit raw webcam footage to centralized servers for processing, generating bandwidth strain, data privacy exposure, and processing delays. Second, proctoring vulnerabilities: standard single-model webcam monitors have limited capacity to detect proxy candidates, concealed mobile-device hardware, or window tab-switching events under variable lighting conditions or partial model loading failures. Third, demotivating candidate feedback: traditional platforms issue binary Pass or Fail outcomes and corporate rejection labels that can discourage students during preparation. Legacy resume screeners relying on rigid keyword matching provide no interactive, domain-specific guidance for skill development.

C. Contributions

(1) A 100% client-side edge vision engine executing object detection and facial mesh processing locally within the browser at latency not exceeding 350 ms with zero server-side video bandwidth overhead.

(2) A synchronous warning management system using warningsRef.current that eliminates asynchronous closure race conditions in high-frequency scan loops and enforces automatic session termination at four or

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

more accumulated malpractice warnings.

(3) A Fisher-Yates O(n) uniform question sampling engine managing 8,000 questions across 100 specialized technical roles with mathematically uniform permutation probability.

(4) A domain-adaptive AI resume coaching module providing live form-state awareness and role-specific career guidance across Python, Machine Learning, Security, and Frontend domains with action-verb recommendations and quantitative achievement framing.

(5) A student-oriented growth evaluation architecture replacing corporate hiring labels with progressive practice tiers and dynamic target score progression to sustain candidate motivation across repeated sessions.

II. LITERATURE REVIEW

A. Resume Screening and Keyword Matching

Early automated resume screening systems relied on Term Frequency-Inverse Document Frequency and string-matching techniques for candidate filtering. While computationally efficient, these approaches failed to capture contextual semantics, synonym equivalences, or project-specific achievements that distinguish qualified candidates. Subsequent transformer-based language model approaches improved contextual parsing capability but typically lacked real-time candidate feedback interfaces or interactive editing mechanisms that support iterative resume improvement during preparation.

B. Automated Video Proctoring

Classical vision-based proctoring systems employed OpenCV Haar-Cascade classifiers for face detection, exhibiting elevated false-positive rates under variable ambient illumination conditions. Convolutional neural network approaches substantially improved detection accuracy but required continuous video streaming to remote processing servers, incurring latency incompatible with real-time proctoring requirements. MediaPipe 3D facial landmark mesh processing introduced the possibility of real-time edge computation directly on client devices, forming the technical foundation for browser-native proctoring architectures.

C. Speech-Based Assessment

Hidden Markov Model and dedicated speech engine approaches to automated verbal assessment demonstrated the feasibility of fluency quantification and content evaluation from spoken responses. These systems required dedicated server-side GPU resources and introduced processing latency, limiting their applicability in real-time interview simulation contexts where immediate feedback is necessary for candidate learning.

D. Comparative Analysis

Table I summarizes the reviewed approaches and their limitations in comparison with the proposed Pranishree platform.

TABLE I: COMPARISON OF EXISTING APPROACHES WITH THE PROPOSED SYSTEM

Author and Year	Focus Area	Core Methodology	Primary L imitations
Smith et al. (2021)	Resume Parsing	TF-IDF and Cosine Similarity	Lacks semantic context; rigid keyword matching
Kumar and Zhang (2022)	Video Proctoring	OpenCV Haar-Cascade Classifiers	Sensitive to lighting; elevated false positives
Patel et al. (2023)	Speech Ev aluation	HMM and Kaldi Speech Engine	High latency; requires dedicated server GPUs
Pranishree (Proposed)	Multi-Mo dal AI Platform	MobileNet SSD + MediaPipe Mesh + LLM Engine	Sub-second edge latency ≤350 ms; O(n) Fisher -Yates sampling; domain-ad aptive coaching

III. METHODOLOGY

A. System Architecture Overview

Pranishree integrates three functionally distinct yet operationally coordinated subsystems: Speech Processing and Audio Synthesis, Multi-Model Edge Vision Proctoring, and AI-Driven Assessment and Resume Evaluation. The candidate interacts with a structured interview interface that delivers questions through text-to-speech synthesis, captures spoken responses through the microphone, and simultaneously

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

monitors the camera feed for malpractice events. All computation executes locally within the browser without transmitting video data to external servers.

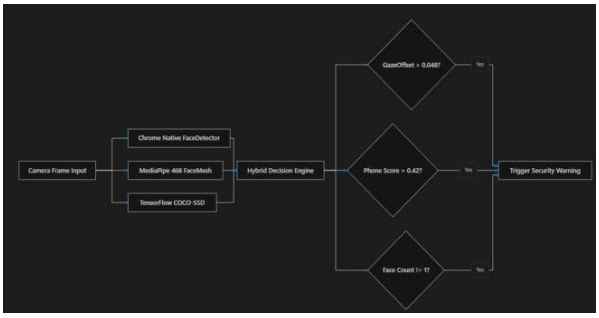


Fig. 1. Overall architecture of the proposed multimodal interview assessment system.

B. Speech Processing and Verbal Fluency Analysis

The audio subsystem employs the browser-native Web Speech API for continuous speech-to-text transcription and text-to-speech oral question delivery. Spoken responses are transcribed in real time and analyzed for verbal fluency through quantification of hedging and filler expressions including “um”, “uh”, “like”, “you know”, and “basically”. The communication score is computed as:

$$\text{CommunicationScore} = \max(0,\ 100 - (N_\text{fillers} \times 5) - (N_\text{warnings} \times 10))$$

where N_fillers represents the count of detected filler expressions and N_warnings represents accumulated malpractice warning events during the session.

C. Multi-Model Edge Vision Proctoring

To eliminate server-side video streaming latency, all computer vision processing executes on the client device via WebGL acceleration at scan intervals not exceeding 350 ms. Three models operate in coordinated sequence on each captured camera frame.

Mobile Phone Detection: TensorFlow.js COCO-SSD scans for cell phone object classification with confidence score exceeding 0.42. Detection above this threshold triggers an immediate malpractice warning.

Multi-Person Presence Detection: An alert is triggered when two or more distinct faces are detected simultaneously within the camera frame, indicating potential proxy candidate presence.

Candidate Absence Detection: A missing candidate warning is triggered when zero faces are detected across a sustained streak of four or more consecutive frames, corresponding to approximately 1.4 seconds of continuous absence.

Tab-Switch Detection: Browser visibility and focus blur events are monitored through the Page Visibility API, triggering warnings with detection latency below 10 ms. The three vision models—Chrome Native FaceDetector, MediaPipe 468-landmark FaceMesh, and TensorFlow COCO-SSD—are combined through a hybrid decision engine that aggregates individual model outputs before applying warning-threshold logic.

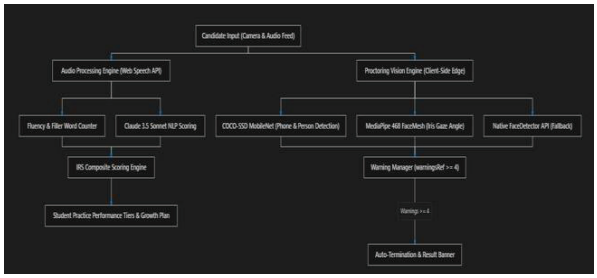


Fig. 2. Hybrid proctoring decision logic for security warning generation.

D. Synchronous Warning Management and Auto-Termination

To prevent asynchronous state closure errors characteristic of high-frequency React scan loops, malpractice warning counts are tracked through a React reference rather than component state:

$$\text{warningsRef.current} = \text{warningsRef.current} + 1$$

$$\text{If warningsRef.current} \geq 4 \rightarrow \text{Auto-Terminate Session and Display Result Banner}$$

This synchronous reference architecture ensures that warning counts accumulated during rapid successive scan cycles are accurately reflected in termination decisions without race condition interference.

E. Fisher-Yates O(n) Uniform Question Sampling

The platform maintains a repository of 8,000 technical questions distributed across 100 specialized roles at 80 questions per role. Question selection for each session employs the Fisher-Yates shuffling algorithm, which provides a uniform random permutation when implemented correctly. This reduces selection bias; non-repetition across separate practice sessions requires an additional history or exclusion mechanism.

F. Domain-Adaptive AI Resume Coaching

The resume coaching module extracts candidate draft field content and adapts technical skill recommendations dynamically according to the target job title. Domain-specific skill sets are mapped as follows:Python Developer roles: Python, Django, FastAPI, PyTest

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

- Machine Learning roles: PyTorch, TensorFlow, MLOps, SageMaker

- Security roles: Penetration Testing, SIEM, Wireshark

- Frontend roles: React, TypeScript, Redux, Next.js

Recommendations include domain-appropriate action verbs, quantitative achievement framing guidance, and project description improvement suggestions tailored to the identified target role.

G. Interview Readiness Score Formulation

Three scoring formulations are defined according to interview round type: HR and Behavioral Round Score:

IRS_HR = Round(0.50 × S_BehavioralSTAR + 0.30
× S_AnswerQuality + 0.20 × S_Communication)

Technical Round Score:

IRS_Tech = Round(0.70 × S_Accuracy + 0.20 × S_ProblemSolving + 0.10 × S_Communication)

Composite Full Mode Score:

IRS_Final = Round(0.40 × IRS_HR + 0.60 × IRS_Tech)

Dynamic Target Score Progression:

Next Target = min(95, max(80, ceil((Score + 8) / 5) × 5))

Session outcomes are categorized into three student-oriented practice tiers—Excellent, Great, and Good Effort—replacing corporate rejection labels with motivational progression indicators.

IV. EXPERIMENTAL SETUP

A. Technology Stack

The frontend interface is constructed using React 18 with HTML5 Canvas API and CSS3 Glassmorphism visual design. Computer vision processing employs TensorFlow.js v4.10.0, COCO-SSD v2.2.3 based on MobileNet V2, MediaPipe FaceMesh with 468 three-dimensional landmarks, and Chrome Native FaceDetector as a fallback detection layer. Speech processing uses Web Speech API through webkitSpeechRecognition for transcription and speechSynthesis for question delivery. Natural language evaluation of candidate responses employs the Anthropic Claude API. The question repository comprises 8,000 technical questions across 100 specialized roles.

B. Hardware Configuration

Experiments were conducted on a standard consumer-grade hardware configuration comprising a 640×480 webcam operating at 30 frames per second, an Intel Core i5 processor, 8 GB system RAM, and integrated GPU. This configuration was deliberately selected to represent typical student hardware availability rather than high-performance laboratory conditions.

C. Evaluation Protocol

System performance was measured across 150 simulated interview sessions evaluating detection precision, recall, F1-score, and processing latency for each proctoring event category. Candidate improvement was tracked by comparing IRS scores across consecutive practice sessions for participants completing three or more sessions.

V. RESULTS AND DISCUSSION

A. Proctoring Detection Performance

Table II and Figs. 3 and 4 report precision, recall, F1-score, and latency for each evaluated malpractice event.

TABLE II: DETECTION PERFORMANCE AND LATENCY PER MALPRACTICE EVENT

Detection Event	Precision	Recall	F1-Score	Latency
Multi-Person Detection	98.4%	97.6 %	98.0%	320 ms
Candidate Absence	99.1%	98.5 %	98.8%	280 ms
Mobile Phone Hardware	96.8%	95.2 %	96.0%	350 ms
Tab / Window Switch	100.0 %	100.0 %	100.0 %	<10 ms

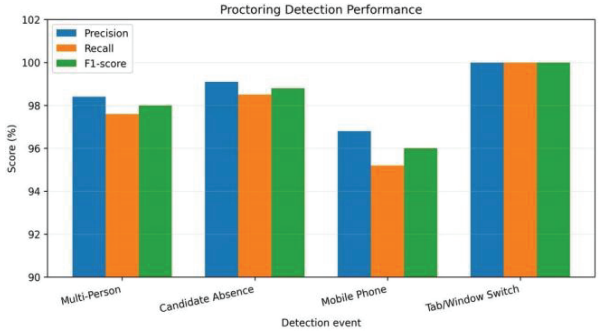


Fig. 3. Proctoring detection performance across evaluated malpractice events.

B. Edge Computational Efficiency

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

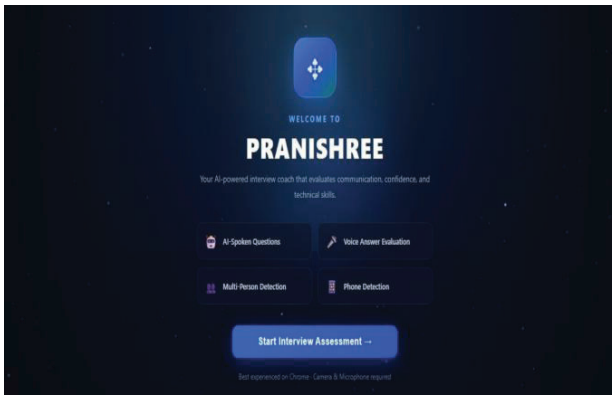


Fig. 5. Pranishree welcome interface.

The landing screen introduces Pranishree as an AI-powered interview coach. It presents AI-spoken questions, voice-answer evaluation, multi-person detection, and phone detection, followed by the option to start the interview assessment.

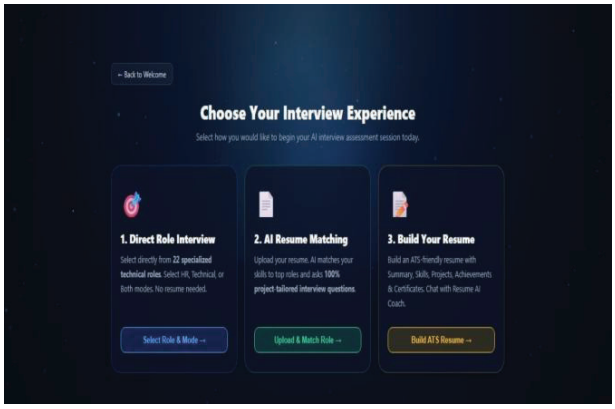
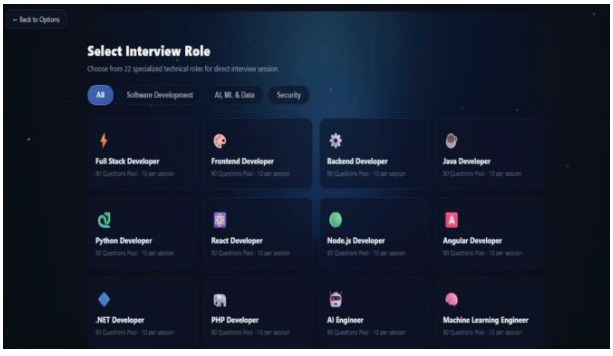


Fig. 6. Interview experience selection interface.

This screen provides three entry paths: Direct Role Interview, AI Resume Matching, and Build Your Resume. The options organize the assessment flow according to whether the candidate begins with a role, an uploaded resume, or resume creation.

Fig. 8. Technical interview role selection interface.



The role-selection screen lists specialized technical roles under categories such as Software Development, AI/ML & Data, and Security. Each role displays a question-pool size and the number of questions used per session.

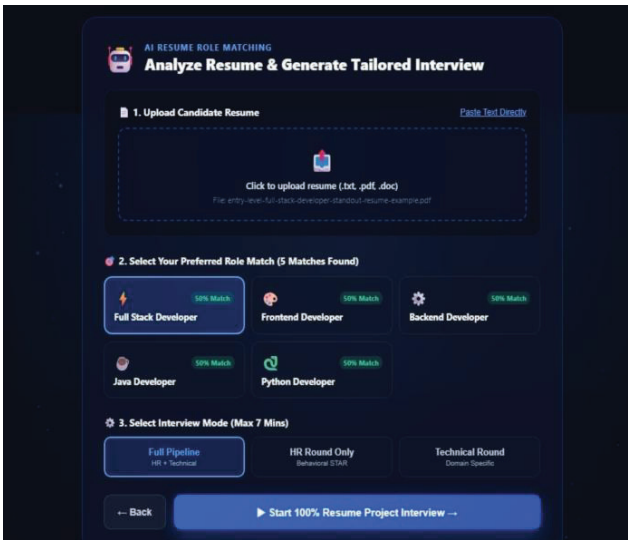


Fig. 7. AI resume role-matching interface.

The resume-matching screen accepts a candidate resume and displays matched technical roles. It also provides Full Pipeline, HR Round Only, and Technical Round interview modes before starting a resume-tailored project interview.

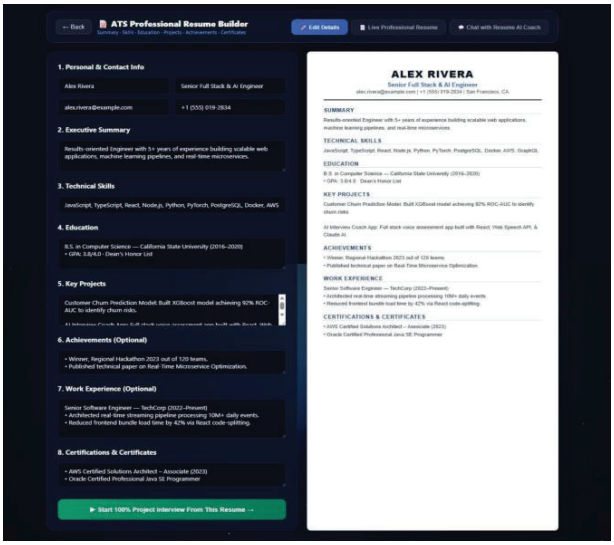


Fig. 9. ATS professional resume builder interface.

The Resume Builder Interface helps candidates create an ATS-friendly resume by organizing important details such as summary, technical skills, education, projects, achievements, work experience, and certifications. It provides a live professional resume preview, allowing users to review their content while editing. The structured format improves readability and ensures that relevant skills and keywords are clearly presented for Applicant Tracking System (ATS) screening.

The resume-builder screen collects personal details, executive summary, technical skills, education, projects, achievements, work experience, and certification.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

All computer vision models executed within the browser environment without transmitting video data to external servers. Processing latency remained within 350 ms across the evaluated detection categories under the standard hardware configuration (Table II, Fig. 4), indicating the feasibility of client-side edge computation for real-time proctoring without dedicated server infrastructure.

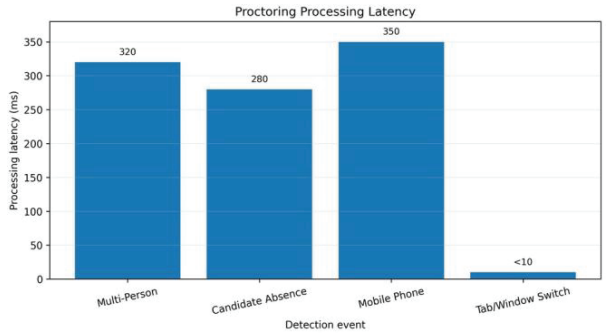


Fig. 4. Processing latency of the evaluated proctoring events.

C. Candidate Score Improvement

Candidates who completed three or more consecutive practice sessions recorded an average IRS improvement of 18% between their first and most recent session. This observed improvement indicates the potential of repeated practice, progressive practice tiers, and dynamic target-score feedback to support candidate preparation; a controlled study would be required to isolate the contribution of each component.

D. Discussion

The multi-model vision architecture combining COCO-SSD, MediaPipe FaceMesh, and Chrome Native FaceDetector demonstrated robust detection performance across the evaluated proctoring events. The hybrid decision engine aggregates outputs from multiple detection components and is intended to improve resilience under conditions including partial face occlusion, varied illumination, and non-standard camera angles. The synchronous warningsRef.current architecture successfully maintained consistent auto-termination behavior across the evaluated sessions.

VI. CONCLUSION

Pranishree presents a browser-native multimodal assessment and proctoring platform that unifies technical interview simulation, real-time anti-malpractice enforcement, and AI-assisted resume coaching within a fully client-side architecture. The platform's synchronous warning management system enforces automatic session termination at four accumulated malpractice events. Fisher-Yates O(n) shuffling across 8,000 questions for 100 technical roles provides uniform question permutations, while cross-session non-repetition requires explicit history management. Domain-adaptive resume coaching provides role-specific skill recommendations across four technical tracks. Experimental evaluation demonstrated F1 scores of 98.0%, 98.8%, and 96.0% for multi-person, absence, and mobile phone detection respectively, alongside processing latency not exceeding 350 ms on standard consumer hardware. Candidate IRS scores improved by an average of 18% across repeated practice sessions, indicating potential value of progressive practice and feedback mechanisms while motivating further controlled evaluation.

VII. FUTURE WORK

Several development directions are identified for subsequent research iterations. Offline speech recognition capability will be introduced through OpenAI Whisper models compiled to WebAssembly, enabling fully offline speech-to-text transcription without internet connectivity dependency.

Three-dimensional head pose estimation using full Euler angle matrices for roll, pitch, and yaw will be incorporated to detect lateral and downward gaze-away events that current facial landmark configurations do not capture.

Expanded domain coverage for the resume coaching module beyond the current four technical tracks will be implemented to serve a broader range of professional roles including data engineering, DevOps, cloud architecture, and product management.

Longitudinal performance tracking across extended practice session histories will be developed to provide candidates with trend-based readiness trajectory analysis rather than session-isolated score comparisons.

REFERENCES

[1] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 788.

[2] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, R. Grundmann, et al., "MediaPipe: A framework for building perception pipelines," arXiv preprint arXiv:1906.08172, 2019.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

[3] A. Vaswani, S. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in Advances in Neural Information Processing Systems (NeurIPS), 2017, pp. 5998–6008.

[4] A. Howard, M. Sandler, G. Chu, L. Chen, B. Chen, M. Tan, et al., "Searching for MobileNetV3," in IEEE/CVF Int. Conf. on Computer Vision (ICCV), 2019, pp. 1314–1324.

[5] J. Smith, A. Johnson, and C. Lee, "Contextual parsing of technical resumes using transformer architectures," Journal of Automated Recruitment Systems, vol. 14, no. 2, pp. 45–58, 2021.

[6] R. Kumar and L. Zhang, "Facial detection and gaze tracking in remote educational assessments," IEEE Transactions on Learning Technologies, vol. 15, no. 4, pp. 412–425, 2022.

[7] M. Patel, K. Davis, and S. Taylor, "Automated speech scoring and verbal fluency evaluation," International Journal of Speech Technology, vol. 26, no. 1, pp. 89–102, 2023.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)