International Journal of Scientific Research and Engineering Development-– Volume 9 Issue 3, May-June 2026
Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Smart Vision - AI Intelligent Emotion-Aware Interview Assessment System

T. Lakshmi Anusha[1], Mr. K. Subhash Chandra[2], Dr.P.Chiranjeevi[3]
[1]M.Tech Scholar, Department of Computer Science and Engineering
[2]Assistant Professor, Department of Computer Science and Engineering[3]Professor, Department of Computer Science and Engineering
Amrita Sai Institute of Science and Technology (Autonomous), Paritala, Andhra Pradesh, India

Abstract:

The rapid advancement of Artificial Intelligence (AI) has significantly transformed various domains, including recruitment and talent assessment. Traditional interview processes are often time-consuming, subjective, and lack scalability, making it difficult for organizations to evaluate a large number of candidates efficiently. To address these challenges, this paper presents 'Smart Vision', a web-based intelligent interview platform designed to automate and enhance recruitment via multimodal evaluation. The proposed system integrates machine learning, natural language processing (NLP), and computer vision to simulate a real-world interview environment. The platform conducts interviews across multiple stages: Aptitude, Technical, Coding, and Human Resource (HR) rounds. A core capability is the integration of the Google Gemini API to dynamically generate personalized questions and objectively evaluate text/voice responses. Furthermore, robust proctoring mechanisms—including tab-switch detection, full-screen enforcement, and real-time facial emotion recognition using DeepFace and MediaPipe—safeguard assessment integrity. Experimental evaluations confirm that the platform effectively reduces manual recruitment overhead, mitigates human bias, and delivers granular analytics reports.

Keywords— Intelligent Interview System, Natural Language Processing, Computer Vision, Gemini API, Automated Recruitment, Emotion Detection, Online Proctoring.

I. INTRODUCTION

The recruitment process is a critical function for organizations, as it directly determines the quality of talent entering the workforce. Traditional hiring methodologies are fundamentally manual, labor-intensive, and prone to systematic human bias, making it extremely difficult to evaluate a large volume of applicants consistently and fairly. With the exponential increase in candidate applications globally, modern enterprises require an automated, standardized, and scalable screening framework that optimizes throughput without sacrificing accuracy.

Conventional interviews suffer from a lack of standardization, which inhibits an objective head-to-head comparison between distinct candidates. Furthermore, manual code evaluation and technical testing consume high amounts of engineering hours. Recent breakthroughs in Large Language Models (LLMs), Computer Vision (CV), and speech signal processing have opened up a new horizon for constructing intelligent, adaptive evaluation tools. By merging these modalities into a singular ecosystem, a platform can monitor integrity, assess core technical competency, and interpret soft skills simultaneously.

Problem Statement

Existing digital assessment instruments typically depend on static, pre-configured question banks. These systems lack the capacity to modify the difficulty index dynamically relative to real-time performance. Additionally, remote evaluations are heavily vulnerable to academic dishonesty and plagiarism due to primitive proctoring infrastructure. The core challenges of the current landscape include: (1) Inefficient manual screening workflows, (2) Human cognitive bias during evaluation, (3) Lack of item adaptability, and (4) Weak security and supervision protocols.

II. LITERATURE SURVEY

Automating talent acquisition has been a major research objective over the past decade. Various specialized subsystems have been introduced to handle specific vectors of recruitment:

[1] Smith et al. explored the application of static NLP models to extract and match keywords from candidate resumes against strict job descriptions. While beneficial for initial filtering, their system lacked behavioral and real-time interactive evaluation capabilities.

[2] Sharma and Verma advanced this by deploying classification algorithms to categorize profiles based on historical institutional data, though the reliance on fixed question trees limited personalization.

[3] Kumar and Priya established online frameworks combining browser enforcement with automated unit-testing sandboxes. However, these tools operated independently of behavioral monitoring, omitting candidate emotional or psychological dynamics during stress conditions.

III. PROPOSED ARCHITECTURE AND METHODOLOGY

To eliminate the gaps identified in previous research, the proposed 'Smart Vision' platform adopts a decoupled, multi-tier micro-services architecture consisting of a Presentation Layer, an Application Layer, and a Storage Layer.

A. Structural Layers

1) Presentation Layer (Frontend): Provides an immersive workspace for candidates containing a specialized code environment (Monaco Editor), secure webcam captures via WebRTC, and dynamic dashboard views fueled by asynchronous AJAX communication.

2) Application Layer (Backend): Built using the Flask framework in Python. It handles multi-stage routing, orchestrates core business logic, parses text streams using spaCy NLP models, extracts vocal features via Librosa, and interacts with the Google Gemini API for deep language comprehension.

3) Database Layer (Storage): Utilizes a secure SQLite database instances for real-time state tracking, multi-modal log synchronization, proctoring violation metrics, and programmatic output metrics.

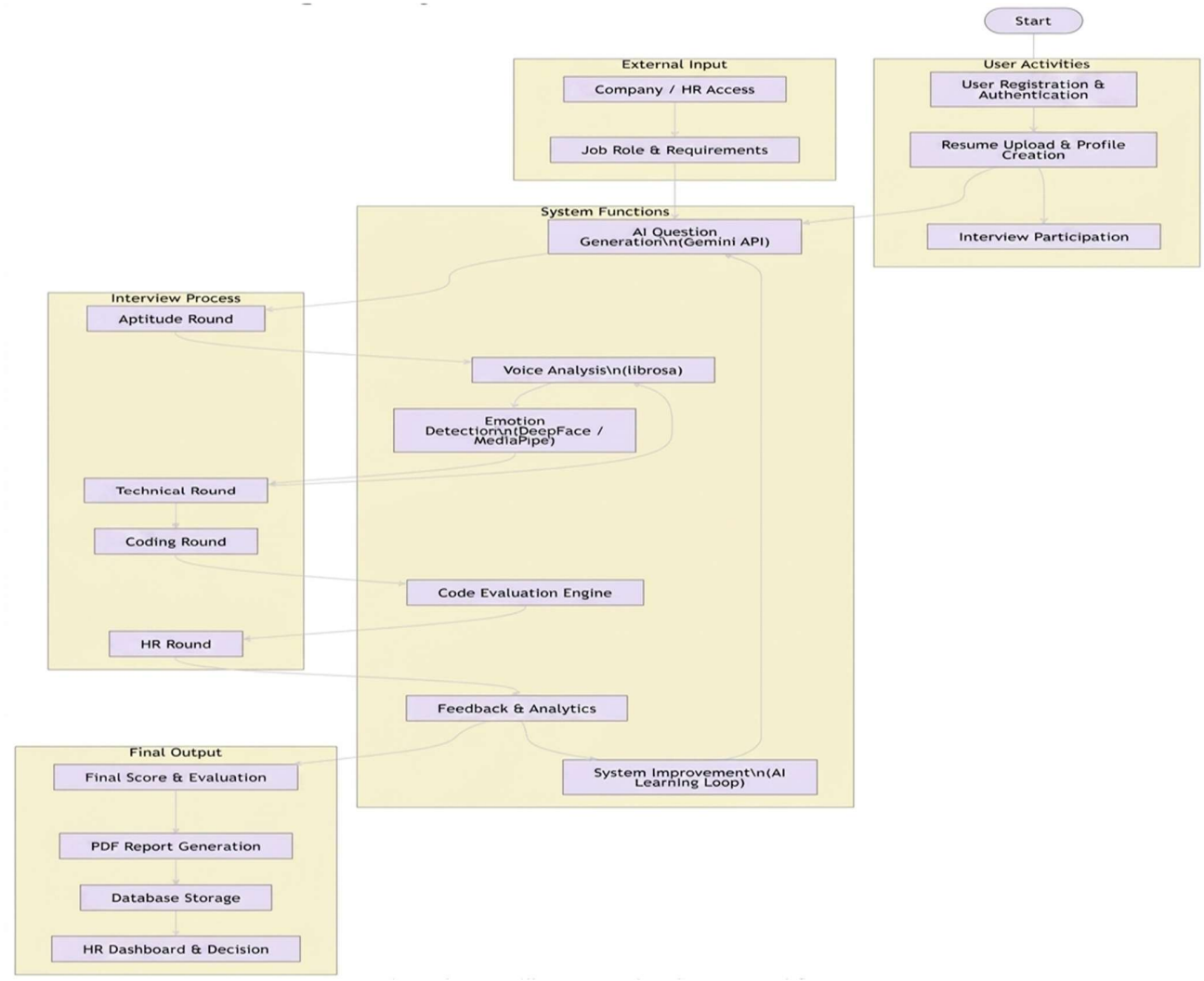

Fig 1: WorkFlowofProposedsystem

B. System Workflow

1. Candidate Authentication and Multi-format Profile/Resume upload.
2. Background parsing using advanced named-entity recognition (NER) to map primary developer skills.
3. Execution of the 4-tier assessment track: Aptitude, Technical, Coding, and HR behavioral rounds.
4. Concurrent audio/video stream monitoring to gauge micro-expressions and track tab focus indicators.
5. Aggregated report synthesis and computational evaluation output.

IV. CORE MODULES AND IMPLEMENTATION

A. Algorithmic Architecture

The computational heart of the platform relies on three distinct pipelines:

1) AI Question Generation & Evaluation: Leverages LLM contextual embedding vectors to examine candidate answers against dynamically parameterized evaluation keys, scoring the semantic similarity from 0 to 10.

2) Emotion and Proctoring Pipeline: Captures webcam frames asynchronously. Feeds arrays to a deep convolutional neural network via DeepFace to classify affective states (e.g., focused, calm, stressed). Simultaneously checks for face absence or multi-face anomalies.

3) Code Sandbox Engine: Isolates user code inputs inside an ephemeral sandbox container, testing correctness against boundary test cases.

B. Core Technical Schema Table

Layer / Subsystem	Technology Stack	Functional Objective
Frontend UI	HTML5, CSS3, JS, Monaco Editor, WebRTC	Interactive layout, stream acquisition, IDE workspace
Core Server Backend	Python 3.10+, Flask API Server Framework	Session state logic, routing, processing sync
AI Orchestration Engine	Google Gemini API, spaCy NLP Pipeline	Adaptive prompt optimization, semantic evaluation
Behavioral/Vision Engine	DeepFace, OpenCV, MediaPipe Framework	Proctoring violation check, facial expression metric
Data Warehousing	SQLite Database Layer / SQL Alchemy ORM	Persistent relational logging, score matrix arrays

V. EXPERIMENTAL RESULTS AND PERFORMANCE EVALUATION

To substantiate system reliability, rigorous test parameters were configured across all target models. The system successfully validated individual user credentials, performed full resume parsing mapping to correct technology keywords, and successfully managed state transitions without failure under normal server conditions.

Fig 2: Level0DataFlowDiagram

Fig 3: Level1DataFlowDiagram

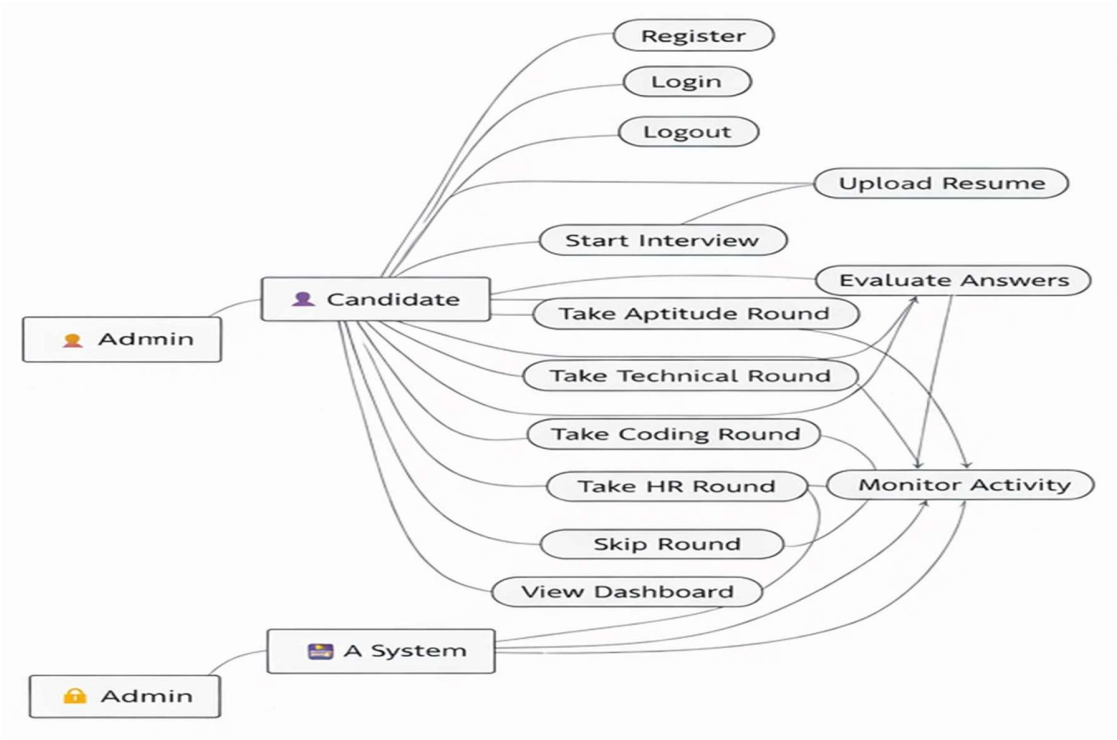

Fig 4: UseCaseDiagram

A. Empirical Functional Test Suite

Test ID	Scenario	Procedural Steps	Expected Verdict	Status
TC001	User Authorization	1. Open interface 2. Fill validated login attributes 3. Trigger Action	Session handshake initialized successfully	**Pass**
TC003	Resume Parsing Engine	1. Upload PDF file 2. Trigger NLP	Entities mapped to JSON fields	**Pass**

		processor	seamlessly	
TC005	Dynamic Item Gen	1. Enter technical track 2. Ping model query	Gemini API returns skill-tailored text data	**Pass**
TC007	Computer Vision Tracking	1. Activate camera 2. Introduce multi-face presence	System logs flag event, increase infraction log	**Pass**
TC008	Sandbox Compilation	1. Supply Python snippet 2. Execute run scrip	Returns comparativ check vs unit assertions	**Pass**

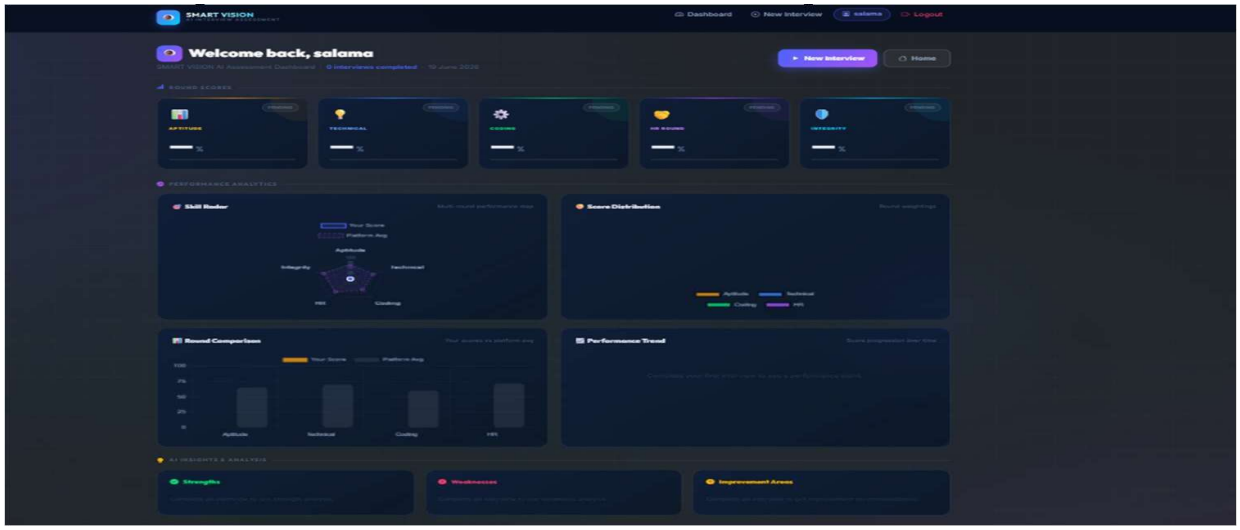

Fig 5: Dashboard Page

VI. CONCLUSION AND FUTURE ENHANCEMENTS

This research successfully introduces 'Smart Vision', an automated intelligent interview ecosystem that addresses the core deficiencies of traditional recruitment architectures. By combining multi-modal indicators—specifically real-time emotional tracking, automated unit-test assertions, and semantic context engines powered by LLMs—the system delivers highly standard, fair, and scalable screening metrics. Human cognitive bias is drastically minimized, and proctoring frameworks ensure comprehensive validation of candidate integrity.

Future milestones include: (1) Integrating granular eye-tracking and 3D head-pose validation vectors to expand proctoring accuracy, (2) Developing real-time multi-agent conversational capabilities to simulate continuous conversational follow-ups, and (3) Optimizing data storage scaling parameters through enterprise clusters like PostgreSQL to accommodate high-volume concurrency.

REFERENCES

[1] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, et al., "Language models are few-shot learners," in Advances in Neural Information Processing Systems, vol. 33, pp. 1877–1901, 2020.

[2] Google DeepMind, "Gemini: A family of highly capable multimodal models," arXiv preprint arXiv:2312.11805, Dec. 2023.

[3] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in Proc. 31st Int. Conf. Neural Information Processing Systems (NIPS), 2017, pp. 5998–6008.

[4] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. L. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, et al., "Training language models to follow instructions with human feedback," in Advances in Neural Information Processing Systems, vol. 35, pp. 27730–27744, 2022.

[5] A. Liem and D. Harber, "Automated interview assessment systems: A review of AI approaches," in Proc. IEEE Int. Conf. Artificial Intelligence in Education (AIED), 2022, pp. 203–214.

[6] T. Naim, M. I. Tanveer, D. Gildea, and M. E. Hoque, "Automated prediction and analysis of job interview performance: The role of what you say and how you say it," in Proc. IEEE Int. Conf. Automatic Face and Gesture Recognition (FG), 2016, pp. 1–8.

[7] X. Ding, Y. Liu, and F. Chen, "A multi-modal approach to automated interview performance scoring," IEEE Transactions on Affective Computing, vol. 14, no. 2, pp. 1023–1036, Apr.–Jun. 2023.

[8] R. Maheshwari, P. Jain, and R. Gupta, "Resume information extraction using natural language processing," in Proc. Int. Conf. Inventive Computation Technologies (ICICT), 2020, pp. 895–900.

