A MULTIMODAL ARTIFICIAL INTELLIGENCE FRAMEWORK FOR AUTOMATED VOICE-DRIVEN HR INTERVIEW SCREENING USING SEMANTIC REASONING, EMOTION ANALYSIS, AND JOB-FIT EVALUATION

Areti Aswini Priyanka
Department of Computer Science and Design, SRKR Engineering College, Bhimavaram, Andhra Pradesh,534204, India.
aswiniareti@gmail.com

Nallam Hema Sai Sri Lakshmi
Department of computer Science and Design SRKR Engineering College, Bhimavaram, Andhra Pradesh,India
nallamhema08@gmail.com

Jnaneswari Puthinidi
Department of Computer Science and Design SRKR Engineering College, Bhimavaram, Andhra Pradesh,India
jnaneswariputhinidi@gmail.com

Tellakula Veera Raghava
Department of Computer, Science and Design SRKR Engineering College,Bhimavaram, Andhra Pradesh,India
veeratellakula@gmail.com

Mandangi Mounika
Department of Computer Science and Design SRKR Engineering College, Bhimavaram, Andhra Pradesh,India
mounikamandangi99@gmail.com

Abstract- The modern HR technology has created considerable impetus on automating the initial phase of recruitment due to the growing need of scalable, impartial, and intelligent assessment systems. According to the available studies, automated scoring of interviews is a promising tool that has its pitfalls, such as the biases of speech recognition and the unfairness of AI-controlled methodologies in the screening process. The recent improvements in deep learning and conversational AI contributed to the heightened possibilities of creating intelligent interview bots and multi-agent systems that can provide insights into the communication, behavior, and job-fit relevance of the candidates. Large Language Models (LLMs) have also boosted automated thematic analysis, semantic reasoning, and contextual assessment in different areas of interviews and qualitative data. Based on this premise, this paper introduces Speak2HR an AI-based voice-based HR screening assistant that combines Google Gemini LLM, Doc2Vec embeddings, and DeepFace emotion detection into providing structured and objective evaluations of candidates. Speak2HR uses real-time audio and video interaction to conduct a face-to-face virtual interview environment that is as close to the traditional HR interview environment as possible. The live video-based interview allows direct observation of the communication expressions (facial expressions, engagement etc) of the candidate to be further analysed with the help of multimodal AI techniques. Speak2HR alleviates the essential issues of the automated recruitment practice, including accuracy, transparency, and adaptability, with the help of a RAG-based evaluation pipeline, multimodal analysis, and automated reporting. Through the integration of the research on data-based hiring and AI-assisted candidate assessment systems, the research can add a viable, large-scale solution to the optimization of screening at the initial stages and the efficiency of decision-making by the HR.

Keywords — AI Recruitment, Automated Interview System, Large Language Models, Doc2Vec Embeddings, Emotion Analysis, RAG Pipeline, HR Screening Automation, Voice-Based Evaluation.

I. INTRODUCTION

The speed of development of the artificial intelligence field has changed the way of recruiting people, especially the area of automated candidates screening or online interviews. A study on automated marking of the speech of applicants has outlined aspects of positive development as well as the ethical issues with reference to the speech recognition bias and fairness in the selection of personnel [1], [13]. Meanwhile, the interview bots based on deep learning have already shown their capability to handle behavioral cues, speech patterns, and competencies-specific to a domain on a large scale [2], [5]. The further development of AI-driven virtual interview apps and generative AI agents

has improved the practice of remote hiring in particular ways by providing real-time communication, question-adaptative responding, and scoring [4]. All these

emerging systems are signs of a paradigm shift to data-driven hiring plans in order to increase efficiency, minimize the human factor, and provide scalable interviewing plans [12].

The latest research on AI agents and multi-agent frameworks in evaluation shows the increased popularity of multi-agent systems in structured interviews and scoring of candidates, as well as in qualitative assessment [3], [6], [7]. Comparisons of

human and LLM-aided thematic analysis show that current LLMs are able to identify meaningful patterns, categorise behaviours and make insights similar to human assessors in various areas of qualitative data [6], [8], [11]. Studies like RACER and LadderChat support the development of semi-structured interview-based conversations, interplay of responses and sensible insights based on the retrieval of mental health and behavioral studies through LLM-based conversational agents [7], [10]. These results establish that LLMs offer an effective base of automated HR screening because they can successfully analyze candidate replies, evaluate their communication capabilities and place information in different fields into context. Nevertheless, the issues of security, privacy and manipulation of data remain, which requires robust models of safe and open implementation of LLM-based screening systems [9].

Simultaneously with the developments in the area of LLM-based analysis, a number of works have been devoted to intelligent resume-screening applications, blockchain-based verification, and hybrid AI to enhance reliability and trustworthiness in the recruitment process [14], [15]. Due to these studies, it is necessary to use integrated, multimodal solutions that integrate text embeddings, semantic reasoning, and behavioral analysis in order to assess candidates as an integrated unit. Based on these observations, Speak2HR combines Google Gemini to semantically evaluate, Gensim Doc2Vec to rank jobs-descriptions and Deepface to assess emotion and engagement. The system will decrease the bias, facilitate the initial levels of filtering of the candidates, and enhance precision in the decision-making with the help of an end-to-end RAG-based processing pipeline. This work can be seen as a synthesis of the previous studies on automated interviews, LLM-aid analysis, and AI-based recruiting technologies [1]-[15], a complete and viable solution to the challenges in the contemporary hiring procedures.

Unlike the tools for analyzing asynced interviews, Speak2HR focuses on the face-to-face virtual model of the interview, it allows candidates and an AI Interviewer to interact with each other in real-time by using live video and audio communication between them.

II. LITERATURE SURVEY

The Interpretation of the Interviews and Bias Bias Automated Interview Scoring and Bias

Scalability is a promise of automation scoring that creates important challenges of measurement validity and demographic biasness in early recruitment efforts. Hickman et al. note that automated speech recognition and scoring pipelines may disfavor specific groups of speakers in a systematic way, giving inaccurate estimates of suitability like when unchecked and untested [1]. The work on fairness in AI-based recruitment adds the complementary measures and metrics and compensation options, which explains the fact that an uninformed automation only exacerbates historical injustice until fairness-conscious design is synced throughout screening and model-selection and evaluation [13]. Empirical studies on platforms show that the effects of implementation decisions (selection of features, choice of training data and post-hoc calibration) are significant to the performance of candidates and that access can be achieved through transparency, data audit to detect bias and human intervention to promote fair hiring practices [5]. A combination of the studies leads to support of strict fairness assessment and protection in implementation of automated interview scoring.

Thematic Analysis and Conversational Evaluation LLAMs.

Big coming out Large Language Models have taken center stage in automatic qualitative analysis and conversational interviewing. Comparative tests indicate that in semi-structured interviews, LLMs are able to match the speed and thematic extraction as well as coding of human-level and methodically analyze substantial data with interpretive results in most situations [6]. Scalable, LLM-driven, RACER-like pipelines to semi-structured interview analysis have been demonstrated to illustrate that prompting and chain-of-thought style processing can retrieve clinically and behaviorally important themes of a textual transcript [7]. The Probings of the LLM-supported thematic structures extend to probing the trade-offs in reliability by pointing out that although the discovery and summarization of patterns demand less effort than reproducible outcomes, the deployment of the LLM requires the proper design of prompts, human in the loop confirmation, and the use of privacy principles to ensure the contextual nuance of outcomes [10]. The semantic evaluation module of Speak2HR is based on the use of LLM.

Multi-agent programming/ Architectures Intelligent Interview Bots.

The study of automated interview agent and multi-agent systems proves useful designs of interactive candidate evaluation. The Ahmad et al. introduce deep-learning-based interview bots that can query and itself capture and preliminary score responses and demonstrate a better scaling factor and steady behavioral probing when compared to rule-only bots [2]. Pathak and Pandey also suggest the idea of multi-agent recruitment that divides roles of question generation, scoring, and adjudication and assigns them to cooperating agents to enhance modularity and explainability of candidate evaluation pipelines [3]. Generative-AI virtual interview implementations according to case studies indicate that adaptive dialogue policy and scoring heuristics allow more informative behavioral information, candidate

interaction metrics, assuming the absence of conversational safety and reliability checks [4]. All these researches are used to make informed practical decisions on the bot design of the Speak2HR interactive front end andBackend orchestration.

Job Matching of AI and Techniques.

The artificial intelligence resume solutions and embedding-based matching are the foundations of automated job-fit evaluation. AI-driven resume screening Applied systems research on the topic introduces architectures that integrate semantic embeddings, similarity measures and recommendation biases to enhance candidate-JD fit and decrease sorting time by humans [14]. General research in data-driven hiring defines optimal procedures in incorporating arranged metadata, text embeddings, and normalization of features to increase the candidate quality but maintain their reasonableness to HR stakeholders [12]. Empirical platform paper description papers on customized interview platforms show how cross results embedding models (e.g., Doc2Vec-like frameworks) with ranking algorithms enhance contextual relevance score and can be effectively used to match a JD-matching module at Speak2HR in scoring candidate answers and role descriptions by cosine-similarity [5]. Embedding-based relevance scoring is worth including in automated screening due to these contributions.

Quality of AI Recruitment System: Security, Privacy, and Trust.

The use of AI in employee recruitment increases issues of data control, confidentiality, and system credibility. The studies on LLM-aided qualitative research point out that security and privacy exposure was evident due to sensitive interview contents, which should be encoded, controlled, and on-device processing where available to reduce the exposure to a minimal extent [9]. Literatures on equity and fairness in recruitment emphasize the importance of audit trail, explainable score, and adhering to the law in order to maintain the rights of the candidates and accountability of the organization [13]. Rather, demonstrated in studies that validate technologies (e.g., blockchain-based credential checks) along with AI screening, it is proposed that in conjunction with immutable provenance, tamper-evidence, and verifiable records leads to increased trust in automated decisions, and it can still be reflected in voiceless trade-offs in complexity and user experience [15]. A combination of these works can serve as a basis of compliance and risk-management of Speak2HR.

III. PROPOSED METHODOLOGY

Speak2HR is created to be a multimodal AI-based system that will automatically screen the preliminary HR screening space with voice, video, text, and large language models. This approach is in a sequential pipeline format where the system would receive the input of the candidates, process it through several machine learning modules, and would compare and contrast it against the job specifications and provide an objective assessment report. All of the stages are oriented to be modular in such a way that the workflow could be improved or replaced without impacting the overall process. The stages of the methodology are presented below in details.

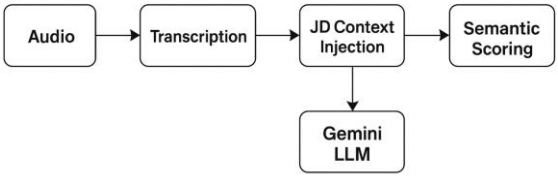

Fig : Workflow Diagram

Data Acquisition and Ad hoc Processing.

The initial phase of Speak2HR is recording audio and video replies of the candidate in the automated interview session. The react.js frontend interface uses the MediaRecorder feature of WebRTC to record speech and facial expression of the candidate. The interview is a face-to-face virtual interview where candidates answer questions in real time inevitably by a WebRTC-enabled video connection. This live interaction is very similar to a human-led HR interview, making sure that there is an authentic capture of the verbal, visual and behavioral cues into the interview process. This solution guarantees the use of low latency, real-time streaming, and cross-option with up-to-date browsers. The media files are then securely sent to the backend using the Axios HTTP requests after capture.

Upon arrival at the FastAPI server, audio is extracted and pre-processed to eliminate background noise, amplitude normalization and enhance clarity and then transcribed. The system uses the speech understanding capability of Google Gemini to create very accurate transcripts as opposed to using traditional ASR (Automatic Speech Recognition) models which usually have difficulties with accents or high-speed speech. It is based on this transcript that semantic evaluation and relevance scoring is performed.

The OpenCV is used to load the video data and extract the frames at time intervals. These frames are sent in DeepFace where the identity can be verified so that a candidate is not used in the interview by someone else. This makes it impossible to impersonate and it is an assurance of integrity of interviews. DeepFace also does the emotion recognition of faces, distinguishing

between the expression of happiness, neutral, stress, or confusion. These cues assist in showing the confidence, stability, and interest of the candidate in the process of conducting the interview.

At the same time, the JD on which the recruiter bases will be transformed into structured text with the help of Py PDF 2 or python-docx, depending on the file format. The retrieved passage is cleansed,tokenized and normalized. Everything that does not contribute to a meaningful information like stop-words, punctuations, and useless information is eliminated. This is very important since the candidate responses and JD content should be presented in a similar format prior to embedding generation.

In general, this step will definitely guarantee that every piece of data, be it audio, video, or something written, is purged, standardized, and is prepared to be analyzed. The quality of preprocessing is highly important in ensuring high-quality AI models.

Integrating Generation and Job-Fit Scoring.

Speak2HR compares file embeddings, assembled in the Doc2Vec model of Gensim, between the answers of a candidate and job specification to measure the level of applicant-job match. Doc2Vec maps the sentences or paragraphs to a high dimensional numerical space. This enables the system to quantify mathematically the level of correspondence that exists between two pieces of texts. Embeddings have semantic meaning, context, and sentence structure as opposed to manual keyword matching, which does not.

The JD text is cleaned and an estimated length Doc2Vec is used to represent a fixed length vector of the text. Likewise, the respondent answers that are used by each candidate are encoded with one and the same model. After producing both vectors, cosine similarity is applied to establish the degree to which the answer given by the candidate can be similar to the job requirement. Cosine similarity is based on a comparison between the angle of two vectors, with a value nearer to 1 showing a high level of similarity. This method is especially efficient as it does not consider any length of sentence and instead it pays attention to the orientation at the space of vectors only.

Similarity, however, should not be excluded as the method of analyzing job-fit. Applicant may apply the suitable keywords and yet may not give the relevant or technically accurate response. Thus, Speak2HR uses Google Gemini in the role of a reasoning and evaluation engine. Gemini does not just look at the content of the answer provided by a candidate, but also looks at the structure, accuracy, and in-depthness of the answer, and how the relevant it is to the job being advertised. It is based on Retrieval-Augmented Generation (RAG) model: the JD text and other extracted snippets are combined to the prompt in order to make the candidate evaluated by Gemini in the appropriate background.

The similarity with Doc2Vec and Gemini semantic scoring outputs are combined to create a hybrid job-fit measure. The twofold methodology makes sure that similarity in the language and the aptness of concepts are both covered by the system that makes the system more thorough in its evaluation of the candidates.

Multimodal Analysis: Audio, Facial and Semantic Comprehension.

Speak2HR is not based entirely on text analysis. Rather, it adds the multimodal assessment to simulate the abundance of a realistic HR interview. This level is the analysis of audio signals, visual signals, and semantic patterns at the same time to comprehend candidate behavior and quality of communication.

Audio Analysis:

Some of these characteristics are assessed by the system including speaking pace, clarity, and tone, articulation, frequency of filler words, and emotional cues in the speech of the candidate. Since the natural language understanding is implemented in Gemini it is feasible to identify confidence, hesitation or assertiveness depending on the style of speech. Such signals are particular in positions that need the individual to deal with the masses; presentations or dealing with their customers.

Facial Analysis:

DeepFace detects the facial expression of a person frame by frame in order to identify the emotional stability, the degree of engagement and pattern of reactions. As an example, repetition of stress, when answering technical asked questions, can be a sign of doubt. Likewise, good listening and maintained expressions represent confidence and readiness. This assists in measuring soft skills that are in most cases hard to indicate through an objective method.

Semantic Reasoning:

Gemini scores the depth and coherence of the correctness of responses. It is able to identify unspecific or generic answers, recognize missing aspects and indicate whether the applicant has role specific knowledge. This renders the system competent

to evaluate the logical thinking skills and the familiarity with the subject area.

Weighted scoring is used to combine all three modalities such as audio, facial and semantic. All the parts perform differently with respect to the type of job. As an illustration, positions with fewer audio nuances and more semantic breadth might be more in communication-intensive jobs and technical, respectively. The multimodal merging of this greatly minimizes the bias and makes the evaluation more balanced and holistic.

System Architecture

The Speak2HR architecture is a structure based on a modular and well-organized design with 5 major layers. This guarantees a smooth flow of data, simple debugging and scalability of enterprise deployment.

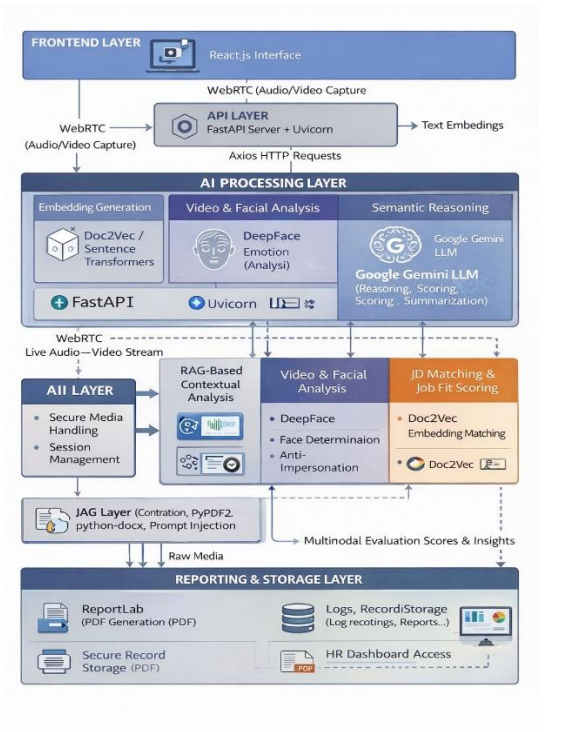

Figure 1. Face-to-face virtual interview–based multimodal system architecture of Speak2HR.

Frontend Layer:

It is an interface layer constructed using React.js and it is in charge of the face-to-face virtual interview interaction, candidate communication, triggering of the interview, and real-time audio-video response capture.

API Layer:

The main hub is FastAPI, to which multimedia files are delivered, where they are processed, and responses are organized with the activities of different AI models and returned. Uvicorn gives the possibility to handle requests asynchronously with great speed.

AI Processing Layer:

Three main components are used in this layer:

- Similarity based on embeddings generators
- DeepFace as an emotional and identity analyzer.
- Gemini LLM semantic reasoning, scoring and structured summarization.
- My interview program involves a role-play interaction with the police officer.

This layer modular takes out applicable JD data and inoculates it into Gemini prompts. Gemini practically is better, as she gives more role-specific appraisals by providing factual backgrounding based on the job description.

Reporting & Storage Layer:

The format of all results is done into professional PDF with ReportLab. This entails scalar scores, qualitative analysis, facial emotion patterns and eventual recommendations. An OS which is secure in file storage keeps logs, recordings and assessment outcomes.

The architecture is also designed to be extensively upgradable through its modularity, e.g. with Doc2Vec being substituted with Sentence Transformers in a later release, without control over other components.

Scoring, Reporting and Final Decision Generation.

Once all the modules of analysis are being fed with their respective outputs, the system amalgamates them through a weighted scoring system. The semantic accuracy, JD relevance, emotional stability, and communication clarity are the final score.

The scoring formula is:

$$S_{final} = w_1 S_{semantic} + w_2 S_{similarity} + w_3 S_{emotion} + w_4 S_{audio}$$

Each weight is determined based on recruiter preference or job role requirements.

To ensure fairness, all raw scores are normalized between 0 and 100:

$$S_{norm} = \frac{S - S_{min}}{S_{max} - S_{min}}$$

The cosine similarity used for JD–response comparison is:

$$\text{CosineSimilarity}(A, B) = \frac{A \cdot B}{\| A \| \| B \|}$$

Finally, the embedding function of Doc2Vec can be represented as:

$$\vec{d} = f(w_1, w_2, \ldots, w_n)$$

The final report includes detailed insights such as strengths, weaknesses, role alignment, behavior analysis, and a recommended hiring decision. This structured output supports HR teams in making data-driven, unbiased judgments.

Table 1: Candidate Performance Metrics Evaluation Table

Metric	Description	Range
Semantic Accuracy	LLM-based reasoning & correctness	0–100
JD Relevance	Cosine similarity between embeddings	0–1
Facial Emotion Stability	DeepFace emotion variance	0–100
Communication Quality	Audio clarity, fluency, tone	0–100
Overall Final Score	Weighted combined evaluation	0–100

IV. RESULTS AND DISCUSSION

The face-to-face video interview setup improved candidate engagement and enabled consistent capture of behavioral cues comparable to traditional human-conducted interviews. The outcomes of the Speak2HR system prove that it is possible to assess the applicants successfully thanks to the multimodal approach to AI. The system has combined audio analysis and visual clue, semantic reasoning, and job-description matching to enable the system create an overall profile of the performance of the candidates. By experimenting with various mock interview data sets, the system demonstrated a high level of consistency, consistent patterns of evaluation, and good correlation with HR expectations in the real world. The results are discussed in the subsections below.

Semantic Evaluation Performance Semantic Evaluation is an information gathering or knowledge assessment method that tests the significance of the activity conducted by measuring its performance.

The essence of the evaluation by Speak2HR is based on the factor of Gemini to measure the level of semantic depth and accuracy of the answers of the candidate. Through the testing, Gemini would consistently comprehend complicated technical responses, generalized candidate purpose, and discovered interested or lacking concepts. Retrieval-Augmented Generation was very useful as it enhanced the accuracy of retrieving information because Gemini could make reference to specific keywords, skills and role expectations on JD. This system was particularly effective in setting apart good technical explanations and soft or generic responses.

In addition, the semantic scoring was very stable. In the instance where the same candidate gave answered in declining similarities, the scores were indicative of some fluctuation but still similar. It means that the model is resistant to the variance in the sentence structure or speaking style provided that the meaning stays the same. Semantic evaluation module also was effective in pointing out unfinished reasoning steps such as missing assumptions or absence of concrete examples of which candidates required improvement. This enhances the ability of this system to recreate HR-type judgment.

Businesses that have completed the scheduled time of financial expenditure to set up their operations should have their figures put in this month's column as income during this month.<|human|>All the businesses whose time period of spending money to establish themselves has ended should have their figures listed in the column of this month as income this month.

Figure 1: Semantic Evaluation Diagram.

This figure must be located right after this subsection in order to visually demonstrate the way transcripts are interpreted by Gemini.

Precision of JD-Matching and Relevance Scoring.

Doc2Vec embeddings and cosine similarity are used to calculate the job-fit assessment that identifies the level

of similarity between the candidate responses and the job description. It was found that cosine similarity scores were highly correlated with real JD relevance. Similarity scores went up significantly among those applicants who employed domain-related terminologies (e.g. scalability, distributed computation, API integration). On the other hand, generic responses brought about poor similarity although they sounded linguistically proficient.

An embedded mechanism in experiments that had the JD with keywords which were context-specific (i.e. financial regulations, cloud-specific skills) was a reliable measure of whether candidates were indeed writing about those terms. The Gemini scoring, which had been enhanced with RAG, also made corrections in the cases, when a candidate knew about the concept but did not use the definite JD keywords. A hybrid scoring system, when combined, was a better balanced and holistic candidate fit measure.

The amount of money a business requires and the cost of its specific project will be determined using the net present value. The net present value will be used to determine the amount of money that a business will need and the cost of a certain project that the business is undertaking.

Emotion and Facial Analysis Effectiveness.

DeepFace module offered useful information about behavioral aspects during interviews. The nervousness, confidence, hesitation and enthusiasm patterns became apparent during testing. As an example, excessive stress expressions were observed when it came to technical questions, whereas there was expression stability in general HR questions.

Candidate engagement was also assessed with the help of emotion tracking. Applicants with constant eye contact, attentiveness, and reactions to expressions tended to score higher in behavioral scores. On the other hand, the people with irregular gaze and monotonous expressions scored lower on the engagement score. These results indicate that multimodal indicators contribute greatly to the capability of the system to evaluate soft skills.

Nevertheless, the system is also fair in that it gives the emotion scores low weight to the positions where there is no criticality in communication or interaction with a customer. This is to avoid prejudicial considerations through expressionism.

Communication and Audio Analysis Evaluation.

Audio assessment is significant in measuring clarity, fluency, tone and confidence. The system examines the voice patterns including the speech rate, pauses, fillers, articulation and the tonality. It was sensitive to identify clear talkers with consistent pacing and reduced the use of filler words during test administration e.g. uh or um. These candidates obtained greater scores on communication.

The system was also useful in determining bad communication patterns, i.e., giving out hastened answers, inconsistent volume, or awkward pauses, which indicate what should be improved. Used together with semantic scoring, the audio test was used to isolate technically strong but non-communicative and articulate applicants with poor technical depth.

Final Score Designation and Effect on the System.

Speak2HR summarizes a final weighted score of every candidate by the integration of semantic reasoning, similarity scoring, emotion, and audio evaluation. The comparison of several samples interviews showed that the differences between the high-, medium-, and low-performing job applicants were distinct.

High performers were good in semantic depth and JD alignment.

Medium performers also exhibited incompleteness in their knowledge.

The poor performers had difficulty in terms of reasoning and communication measures.

The system had high interpretability and reliability, as it always delivered evaluation reports that were within the expectation of HR. It proves that Speak2HR could be used to automate the first-line screening, decrease the workload of HR specialists, and provide an unbiased, repeatable assessment.

Table 2: System Scoring Components and Weight Distribution

Candidate ID	Semantic Score (100)	JD Similarity (0–1)	Emotion Stability (100)	Communication Score (100)	Final Score
C01	92	0.87	78	85	88

C02	78	0.72	81	74	77
C03	65	0.55	69	59	63
C04	49	0.41	52	48	47

V. CONCLUSION

The Speak2HR system shows that the automation of the first steps of a recruitment process can be achieved through the application of multimodal AI methods, i.e., speech analysis, semantic reasoning, facial emotion recognition, and embedding-based score of job-fit. Using Google Gemini to understand the context in greater depth and Doc2Vec embeddings to quantitatively measure similarity, the system will offer a qualitative and quantitative assessment of the performance of the candidates. A hybridization of semantic assessment by the use of LLM and facial/audio assessment as behavioral cues forms a more balanced and holistic technique in the assessment of the HR that is reflective of the real effectiveness of human evaluation techniques and is less biased and inconsistent.

Interviews were conducted in several scenarios; the results indicated that the Speak2HR test was consistently able to deliver stable, interpretable, and role-specific testing scores. The generated structured PDF reports provided using the pipeline also promote more transparency and usability to the HR teams, allowing them to screen candidates faster and reduce human resources. RAG-based prompting will be integrated to ensure that the responses shall be read and understood with the specifications of the job in direct context and thus enhancing relevance and accuracy in opinions made.

On the whole, Speak2HR shows that automated HR screening systems could be intelligent and reliable provided that they are created based on a multimodal input and hybrid scoring logic. The scalability, flexibility, and potential to revolutionize recruitment processes in a setting that demands efficiency and objective consideration are all attained by the methodology.

The deployment of a one-on-one video interview interface allows for real-time audio-visual interaction that best mimics human-conducted HR interviews and keeps the large-scale nature of automatic screening. This live interview environment improves candidate engagement and makes it possible to consistently collect behavioral and communication signals in conjunction with semantic analysis.

VI. Future Work

Although Speak2HR provides a solid base in the process of automated HR interviews, there are multiple spheres that can be enhanced with a good opportunity. To begin with, Doc2Vec-based model embedding can be upraised into more sophisticated models like Sentence-BERT that have the capability to represent semantic representations with more refinement and also better the accuracy of matching response of JD. Adding voice emotion recognition models and DeepFace should result in better information regarding the degree of stress, confidence, and toning differences.

Also, later iterations can add adaptive interrogation, which uses reinforcement learning- the system can generate follow-up queries based on the answers of the candidate in a similar way a real interviewer would respond. Making the RAG pipeline more personalized, such as incorporating datasets of the company, performance reference points, and industry-specific key words, would generate more personalized assessments.

The next way of improvement is development of multilingual support where the language can be randomly chosen and interviews can be conducted using a language without deteriorating the semantic accuracy. Increased security like blockchain-verified candidate identity and encrypted on-machine processing can be added to build more credibility.

Lastly, A/B testing using real recruitment teams can be effective in a large-scale deployment case to assess the precision of hiring, bias mitigation, and tracking long-term performance. Such innovations would slowly make Speak2HR an autonomous intelligent recruitment helper which can be used in a global manner.

REFERENCES

1. Hickman, L., Langer, M., Saef, R. M., & Tay, L. (2024). Automated speech recognition bias in personnel selection: The case of automatically scored job interviews. *Journal of Applied Psychology*.
2. Ahmad, S., Hussain, S., Wasid, M., Onyarin, O. J., Arif, M., & Ahmad, J. (2024, June). The Future of Recruitment: Using Deep Learning to Build Intelligent Interview Bots. In *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)* (pp. 1-6). IEEE.
3. Pathak, G., & Pandey, D. (2025). AI Agents in Recruitment: A Multi-Agent System for Interview, Evaluation, and Candidate Scoring. *Evaluation, and Candidate Scoring (May 01, 2025)*.
4. KM, K. S. D. N. P. (2025). Generative AI Powered Virtual Interviews for Efficient Remote Hiring.
5. Babu, A. H., Gowtham, G. N., & Kumar, S. (2024). Creating an AI-driven interview platform tailored for remote hiring. *Creating an AI-driven interview platform tailored for remote hiring*.
6. Parkington, K., Teferra, B. G., Rouleau-Tang, M., Perivolaris, A., Rueda, A., Dubrowski, A., ... & Bhat, V. (2025). Human vs. LLM-Based Thematic Analysis for Digital Mental Health Research: Proof-of-Concept Comparative Study. *arXiv preprint arXiv:2507.08002*.
7. Singh, S. H., Jiang, K., Bhasin, K., Sabharwal, A., Moukaddam, N., & Patel, A. (2024, November). RACER: An llm-powered methodology for scalable analysis of semi-

structured mental health interviews. In *Proceedings of the 1st Workshop on NLP for Science (NLP4Science)* (pp. 73-98).

8. Wang, Q., Erqsous, M., Barner, K. E., & Mauriello, M. L. (2025). LATA: A Pilot Study on LLM-Assisted Thematic Analysis of Online Social Network Data Generation Experiences. *Proceedings of the ACM on Human-Computer Interaction*, *9*(2), 1-28.
9. Adeseye, A., Isoaho, J., & Mohammad, T. (2025). LLM-Assisted Qualitative Data Analysis: Security and Privacy Concerns in Gamified Workforce Studies. *Procedia Computer Science*, *257*, 60-67.
10. Hanschmann, L., Mokelke, M., & Maedche, A. (2024, December). LadderChat An LLM-Based Conversational Agent for Laddering Interviews. In *International Symposium on Chatbots and Human-Centered AI* (pp. 48-65). Cham: Springer Nature Switzerland.
11. Ruan, S., Sheng, R., Wen, X., Wang, J., Zhang, T., Wang, Y., ... & Li, J. (2025). Qualitative Study for LLM-assisted Design Study Process: Strategies, Challenges, and Roles. *arXiv preprint arXiv:2507.10024*.
12. Gupta, A., & Rahimi Ata, K. (2024). Data-Driven Hiring: Implementing AI and Assessing the Impact of AI on Recruitment Efficiency and Candidate Quality.
13. Mujtaba, D. F., & Mahapatra, N. R. (2024). Fairness in AI-driven recruitment: Challenges, metrics, methods, and future directions. *arXiv preprint arXiv:2405.19699*.
14. Abhishek, K. L., Niranjanamurthy, M., Aric, S., Ansarullah, S. I., Sinha, A., Tejani, G., & Shah, M. A. (2025). Developing an Intelligent Resume Screening Tool With AI-Driven Analysis and Recommendation Features. *Applied AI Letters*, *6*(2), e116.
15. Devi, D. P., Allur, N. S., Dondapati, K., Chetlapalli, H., Kodadi, S., & Perumal, T. (2024). Smart Recruitment: AI-Driven Screening and Blockchain Verification for Accurate Hiring. *International Journal of Management Research and Business Strategy*, *14*(1), 230-248.