AI Mock Interview Behavioural Recognition Analyst

Dr.Reshmi B
Dept Of Cse
Ahalia School Of Engineering And Technology
Palakkad
reshmi.b@ahalia.ac.in

Dr.Gunasekaran Subramanian
Dept Of CSE
ahalia school of engineering and technology
palakkad

Alent Eldho
Dept Of CSE
Ahalia School Of Engineering And Technology
Palakkad
alenteldho123@gmail.com

Glorin Jose
Dept Of CSE
Ahalia School Of Engineering And Technology
Palakkad
glorinjose605@gmail.com

Rohit S
Dept Of CSE
Ahalia School Of Engineering And Technology
Palakkad
rohits@gmail.com

Abstract **:** This proposes a comprehensive framework that integrates a emotion detection , facial recognition, and AI-driven mock interview analysis based on the deep learning techniques. The system utilizes Semantic-Emotion Neural Networks (SENN) to understand emotions from text-based data, YOLO and ArcFace for accurate facial recognition, and a combination of Convolutional Neural Networks (CNN) and Natural Language Processing (NLP) to analyze interview answers. Through the incorporation of these advanced models, the presented system attains improved accuracy and real-time performance in various applications.

To measure its efficiency, the system was subjected to a benchmark data tests, displaying more accurate and better performance in relation to state-of-the-art algorithms. The paper also presents an in-depth look at multimodal deep learning methodology in augmenting human behaviour's recognition and measurement in the online space.

Deep learning and artificial intelligence have transformed the field of human-computer interaction, particularly emotion recognition and behavioral measurement. Contemporary neural network models support real-time emotion detection from audio, text, and visuals. Face recognition has also evolved considerably with deep metric learning, which improves classification accuracy and enhances feature representation. By integrating these technologies, the suggested multimodal framework efficiently addresses essential challenges in real-time human behavior analysis, achieving the best possible trade-off between performance and computational efficiency.

I. INTRODUCTION

In the current competitive job market, interviewing is an essential process to gain a job. As the age of video interviews sets in, there are new challenges for candidates, including how to effectively present themselves on camera and deal with non-verbal cues. Conventional mock interviewing systems may n ot touch on these subtleties, and thus candidates become ill-prepared for the complexities of virtual communication. This paper presents a real-time mock interview system with the capability of offering in-depth feedback to the users. The system simulates a real-world interview setting where candidates can practice respo nding to frequently asked questions with actionable feedback on their performance. Through the evaluation of social behaviors such as facial expressions, head gestures, and speech patterns, the system provides personalized suggestions for users to enhan ce their interview skills.

The framework is built on a software architecture that enables precise control of virtual characters' multimodal behaviors. The virtual character, acting as an AI-driven recruiter, responds dynamically to the user's actions, creating an engaging and interactive experience. Key components of the system include a social cue recognition module, a dialogue and scenario manager, a behavior manager, and a 3D rendering environment.

The primary goals of this research are:

- Developing a deep learning model for emotion recognition in text using BiLSTM and CNN architectures.
- Enhancing face recognition accuracy through real-time face detection with YOLO and feature extraction using ArcFace.
- Incorporating speech analysis techniques with CNNs and NLP algorithms to evaluate interview performance.
- Implementing a multimodal AI framework to improve human-computer interaction and behavioral analysis.
-

By integrating these technologies, the system aims to address the gaps in virtual interview preparation,

empowering candidates to present themselves confidently and effectively in digital interview settings

Related Work

A. *Interview Coaching Systems*

Research has developed interview coaching systems for candidates to rehearse and obtain constructive feedback on their interview skills for many years. The initial models were mostly limited to the basic question-answer paradigm, but the emergence of deep learning has resulted in the creation of sophisticated systems that can analyze non-verbal behavior and provide meaningful feedback. One of the studies by Su et al. [1] proposed an interview coaching system using deep learning with dialog state tracking and action selection. Utilizing Long Short-Term Memory (LSTM) networks and Artificial Neural Networks (ANNs), the system generated relevant interview questions and supervised the progress of the candidate seamlessly.

B. *Speech Recognition And Grammar Analysis*

Speech recognition is one of the key features of interview coaching systems because it deals with the computerized processing and analysis of spoken responses. Brown and Wilpon [2] implemented a grammar compiler to improve the recognition of connected speech, increasing the accuracy of the system. Their contribution was the introduction of Grammar Specification Language (GSL) through which users were able to describe particular task grammars and recognize complex ones. This development aided in the creation of advanced speech recognition systems able to cope with a multitude of interviews and scenarios..

C. *Facial Expression Recognition*

Non-verbal communication is a vital part interviewers pay attention to when processing and analyzing facial expressions. Zhang et al. [4] suggested a method for recognizing facial emotions using CNNs alongside image edge detection methods. Their systems not only achieved high level recognition accuracy but also outperformed traditional systems in processing speed. This is important in real-time interview simulations where prompt feedback on participants' faces can help the coaching..

D. Personality Recognition from Video Interviews

Automatic personality identification from video interviews is a nascent yet a vividly evolving field. With the advancements in artificial intelligence, video interviews can be analyzed for “big 5” personality traits recognition with much precision. Suen et al. [3] developed an AI agent for interviews that was capable of recognizing the “big five” personality traits. Their systems integrated asynchronous video interview (AVI) processing and made use of tensor flow AI features extraction and trait classification pipeline.

E. *Emotion Recognition from Text*

Figuring out the underlying emotions in text answers is also important when looking at their interviews. Bárbatar et al. [6] proposed a model called Semantic-Emotion Neural Network (SENN), aimed to capture the semantic meaning of the data as well as the emotional attributes associated with it. In this case analysis, the model was very useful and atypical in recognizing the emotions of the surface and helped to analyze the responses as well as evaluate the emotional state of the participant during the interview.

These advancements enable modern interview coaching systems to go beyond traditional methods providing richer insight and better feedback for candidates to improve their verbal and non-verbal skills and subsequently perform better in interviews

II. PROBLEM STATEMENT

Despite a rise in demand for communication and soft skills in the workplace, candidates still face difficulties with finding proper resources for interview training. Traditions mock interviews tend to be lengthy, relativistically measured, and resource dependent. Alternatives powered by AI offer a more advanced solution that removes subjectivity by basing performance evaluations solely on the analysis of behavioral patterns and speech. Our proposed model aims to eliminate the discrepancy between 'knowing' and 'doing' by creating interview environments and providing intelligent and customized feedback.

.. System Architecture

This automated interface thoroughly processes user input through voice and video recordings using advanced deep learning models and provides an in-depth evaluation using various performance metrics. Users can use the data to reinforce their weakness and hone their interview skills. Providing users with precise measures of their success allows candidates to correlate their performance with theiroutput.

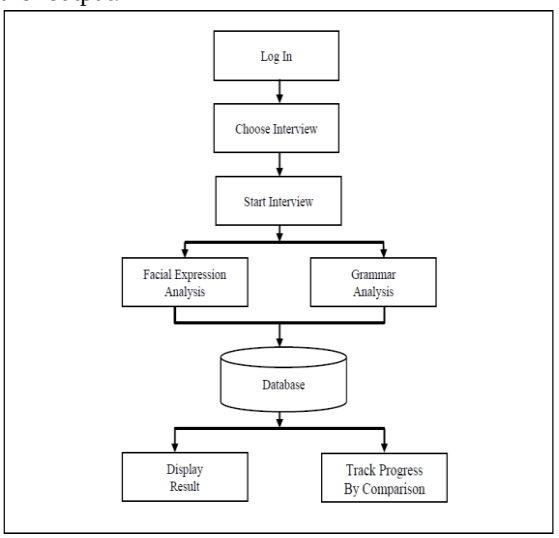
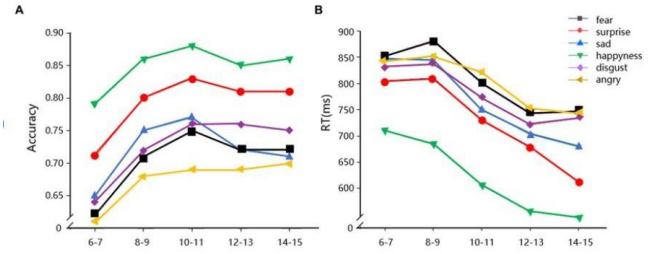


Fig.:1 System Architecture

A. Facial Expression Analysis

During an interview, a candidate's confidence, engagement, and general attitude is best communicated through their facial expressions. Micro-expressions, as well as nodding and eye movement, provide clues into their level of comfort and attention in the background. To identify and interpret these non-verbal signals, our system employs facial recognition as well as sentiment analysis through the use of Convolutional Neural Networks (CNNs) which allow these signals to be detected in real time. Using facial movements and expressions, the model can measure engagement levels and aid the candidates in understanding the impact of their body language on interview performance. As a result, users can improve how they manage their non-verbal signals so that they are portrayed as confident and engaged during business encounters.



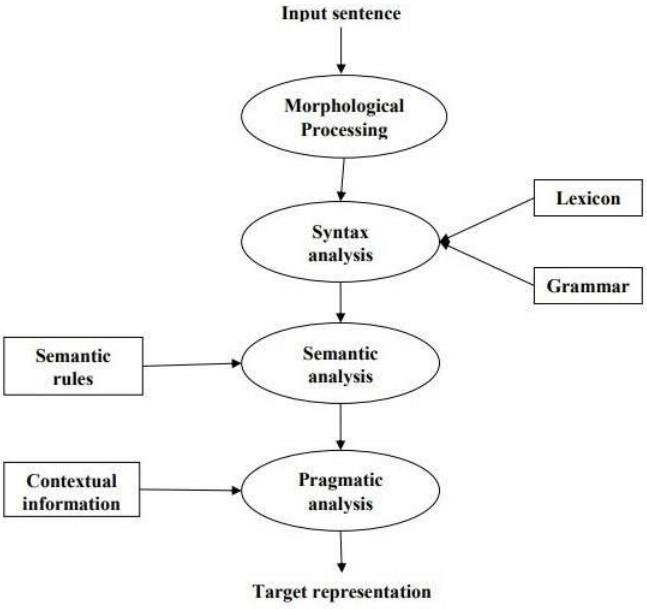
Accuracy (A) and reaction time (B) of facial expression recognition across different age groups.

B. Behavioral Cue Recognition

Behavioral Cue Recognition Nonverbal behaviour is very crucial to the interview outcome because a candidate's body movements can impact their evaluation. The system employs AIassisted behavioral analytics to track head postures and movements and hand gestures as the non-verbal communication behavior being measured. It captures the use of anxious gestures such as rapid movement and lack of adequate body support which can contribute to poor impression. The system helps improve awareness of bodily presence and assists candidates to change posture and body gestures so that they can improve their professional image..

C. Speech and NLP Analysis

This describes the studying of speech and other communication features of Language as well as Speech Patterns and Speech Analysis A successful interview relies heavily on effective communication and our system utilizes Natural Language Processing (NLP) along with Long Short Term Memory (LSTM) models for speech analysis automation. It also checks the grammar, pronunciation and clarity of the speech as well as transcribing the spoken answers. Furthermore, the system evaluates how well a candidate articulates his thoughts by rating the cadence (speaking rate), modulation (volume changes), and articulation (speech clarity). It captures also hesitations and filler utterances which negatively influence the speech fluency and confidence. To make further refinement of the evaluation, the respondents' emotional tone is identified and sentiment analysis is conducted so that comprehensive evaluations with regards to content and delivery are provided.



D. Behavioral Cue Recognition

Behavioral Cue Recognition Nonverbal behaviour is very crucial to the interview outcome because a candidate's body movements can impact their evaluation. The system employs AIassisted behavioral analytics to track head postures and movements and hand gestures as the non-verbal communication behavior being measured. It captures the use of anxious gestures such as rapid movement and lack of adequate body support which can contribute to poor impression. The system helps improve awareness of bodily presence and assists candidates to change posture and body gestures so that they can improve their professional image..

E. Real-Time Feedback Mechanism

In order to ensure the learning process is effective and engaging, instant feedback is provided through imagery and AI analysis. Candidates are given performance scores on several metrics such as clarity of speech, facial expressions, and overall body language. The system sets personalized improvement goals for each user which exercises their communication and presentation skills over a set period of time. With real-time feedback, candidates are able to build self-confidence and improve their interview techniques which prepares them for actual scenarios.

III. IMPLEMENTATION DETAILS

A. *Emotion Recognition*

Emotion recognition is done using Semantic-Emotion Neural Network (SENN) model, which is composed of two deep learning components: BiLSTM based semantic encoder and CNN based emotion encoder. Contextual understanding is improved with the integration of pre-trained word embedding such as Word2Vec and GloVe so that textual responses with emotional semantics are well captured. The model is fine-tuneed with some well known datasets, SemEval and EmoReact, which enables the model to highly accurately recognize a vast array of emotions.

B. *Face Recognition*

For precise face recognition, the system integrates real-time face detection using a YOLO model and robust feature extraction using ArcFace. These models are efficient in locating and recognizing facial features so that they can be matched to stored ones. To improve real-time application performance, TensorRT is deployed to speed up the inference time. The system is trained using Labeled Faces in the Wild (LFW) and VGGFace that provides consistent and accurate facial recognition regardless of the different conditions.

C. *Speech and NLP Analysis*

The system integrates speech and natural language processing (NLP) techniques to evaluate verbal communication skills. Speech-to-text conversion is performed using deep learning-based models, which transcribe spoken responses with high accuracy. The system then analyzes various linguistic aspects, including grammar, pronunciation, speech rate, and clarity. It also detects filler words and hesitations, helping candidates improve their fluency and articulation. Additionally, sentiment analysis is applied to assess the emotional tone of responses, providing insights into how confidently and effectively a candidate communicates.

D. *Mock Interview Analysis*

To give a thorough appraisal concerning a candidate's interview performanceS, the AI-driven system, in this case, captures and evaluates the interview videos in real-time. Using deep learning methods, the head nods, speech, and even the grammar are evaluated. CNN's evaluate facial cues for signs of confidence or nervousness, while the linguistic sign of speech border on achieving coherence is checked via NLP techniques. It was able to provide feedback instantly on the important metrics like the level of engagement, confidence, communication, effectiveness of the interview, among other things, so that the candidates could improve their performance during the interview sessions.

E. *System Integration*

To ensure seamless performance, the system employs a real-time processing pipeline that integrates data from multiple modules. An attention mechanism is used to prioritize significant features from various inputs, such as facial expressions, speech patterns, and textual responses. This enhances the accuracy of decision-making by focusing on the most relevant aspects of the candidate's performance. By combining results from different analytical components, the system delivers a holistic assessment, providing users with meaningful feedback to improve their interview readiness..

IV. EVALUATION

The AI-Based Mock Interview Behavioral Recognition System, which the current study introduces, is one thing that will take a big leap in interview preparedness and assessment. By working with deep learning, natural language processing (NLP), and multimodal analysis of data, this system gives candidates a comprehensive platform to polish their interview skills, grow more self-aware, and advance professionally. A major aspect of it is that it shall give feedback in real-time on what interview performances critically need expressions, speech patterns, and states of emotion. That will make it particularly beneficial for relatively youthful job seekers and college graduates with no access to professional coaching or structured training programs.

This system's main advantages come from its combination of different technologies for the evaluation of the interview. It integrates Convolutional Neural Networks (CNNs) facial expression analysis, Natural Language Processing (NLP) for speech evaluation, and machine learning for personality and emotion recognition, providing full understanding of what the candidate does. This approach allows details to be given to the candidates to ensure they understand their strengths and their weaknesses. For example, the system can identify emotional signs such as anxiety or confident attitudes that may be missed during mock interviews, but are important in real professional environments.

The system's real-time interaction feature immediately stands out the most because users are provided with immediate feedback impressions of their performance. This enhances the interview preparation process as users are able to make instantaneous changes within the system. Various forms such as writing, charts, or even dashboards make feedback more digestable. This is essential for self-improvement focus without interview coaches or mentors.

With all these benefits, the system still has a few limitations. A key problem is the reliability and accuracy of AI models intended for facial expression and emotion recognition. This study claims the use of advanced algorithms, but many elements pose problems such as the surrounding light, video capturing devices, and individual differences of people smiling and frowning. Furthermore, the system will most likely incorrectly identify emotions that are non-verbal, softer or blended because they are the most important in interviews.

Another possible limitation is the system's reliance on fixed criteria to evaluate an interview's success.

Regardless of how well this system guarantees adherence, it poorly handles the intricacies and subtleties of human interaction. It is possible to reach an unfair or entirely wrong assessment due to cultural differences, especially in relation to body language, speech, or any other non-verbal actions

since the system does not capture meaning of various communication styles by different cultures. My AI powered mock interview system solves all these problems and has the potential to transform the traditional approach to teaching interviews and their evaluation to something more flexible. It has an engaging and realistic animation setting, thus bridging the gap between classroom and real-life interview scenarios enabling students to prepare for interviews in more practical ways. In other cases, modifying the approach by measuring users' emotional responses through heart rate and skin conductivity changes might provide better control for advanced guidance after customized feedback is provided. The continued evolution of technology can serve as a motivation for the unemployed by empowering them to believe in themselves, communicate well, and improve their chances of succeeding in complex and highly competitive recruitment processes

REFERENCES

1. Batbaatar, Erdenebileg, et al. "Semantic-Emotion Neural Network for Emotion Recognition from Text." *IEEE Access*, vol. 7, 2019, pp. 111866–111878, ieeexplore.ieee.org/document/8794541, https://doi.org/10.1109/ACCESS.2019.2934529

2. https://www.researchgate.net/publication/380149838_RealTime_Face_Recognition_System_at_the_Edge

3. Zhang, Hongli, et al. "A Face Emotion Recognition Method Using Convolutional Neural Network and Image Edge Computing." *IEEE Access*, 2019, pp. 1–1,https://doi.org/10.1109/

4. Mandal, Rubi, et al. "AI -Based Mock Interview Evaluator: An Emotion and Confidence Classifier Model." *IEEE Xplore*, 1 Feb. 2023, ieeexplore.ieee.org/document/10100589. Accessed 13 July 2023.

5. R. Mandal, P. Lohar, D. Patil, A. Patil and S. Wagh, "AI -Based mock interview evaluator: An emotion and confidence classifier model," *2023 International Conference on Intelligent Systems for Communication, IoT and Security (ICISCoIS)*,

6. Jitendra Madarkar, et al. "Sparse Representation for Face Recognition: A Review Paper." *IET Image Processing*, vol. 15, no. 9, 6 Mar. 2021, pp. 1825–1844, https://doi.org/10.1049/ipr2.12155Siswanto, Joko, et al. "Interview Bot Development with Natural Language Processing and Machine Learning." *International Journal of Technology*, vol. 13, no. 2, 1 Apr. 2022, p. 274, https://doi.org/10.14716/ijtech.v13i2.5018.

7. Vakalopoulou, Maria, et al. "Deep Learning: Basics and Convolutional Neural Networks (CNNs)." *Neuromethods*, 1 Jan. 2023, pp. 77–115, https://doi.org/10.1007/978-1-0716-3195-9_3

8. Dukić, David, and Ana SovicKrzic. "Real-Time Facial Expression Recognition Using Deep Learning with Application in the Active Classroom Environment." *Electronics*, vol. 11, no. 8, 14 Apr. 2022, p. 1240, https://doi.org/10.3390/electronics11081240

9. Shashank Rai, Alisha Miranda, Samiya Jagirdar, Prof. Nidhi Chitalia:Skillup Bot: An AI Driven Mock Interview Platform.

10. AI Based Interview Critique System: Interview Preparation Companion Using Deep Learning.https://www.ijcrt.org/papers/IJCRT2402632

11. Alhanaee, Khawla, et al. "Face Recognition Smart Attendance System Using Deep Transfer Learning." *Procedia Computer Science*, vol. 192, 2021, pp. 4093–4102, https://doi.org/10.1016/j.procs.2021.09.184.

12. Njoku, JudithNkechinyere, et al. "Deep Learning Based Data Fusion Methods for Multimodal Emotion Recognition." *The Journal of Korean Institute of Communications and Information Sciences*, vol. 47, no. 1, 31 Jan. 2022, pp. 79–87, https://doi.org/10.7840/kics.2022.47.1.79

13. S.Yadav, Shital. "EMOTION DETECTION USING DEEP LEARNING ALGORITHM." *Engineering and Technology Journal*, vol. 06, no. 03, 29 Sept. 2021, https://doi.org/10.47191/etj/v6i3.04

14. Prince Awuah Baffour, et al. "A Survey on Deep Learning Algorithms in Facial Emotion Detection and Recognition." *Inform :JurnalIlmiahBidangTeknologiInformasi Dan Komunikasi*, vol. 7, no. 1, 8 June 2022, pp. 24–32, https://doi.org/10.25139/inform.v7i1.4563.